

École Centrale de Nantes

École Doctorale

Sciences et Technologies de l'Information et Mathématiques

Année 2012

Thèse de DOCTORAT

Spécialité : TRAITEMENT DU SIGNAL ET DES IMAGES

Présentée et soutenue publiquement par :

Pauline TROUVÉ

le 10 décembre
à l'ONERA, Palaiseau

TITRE

**Conception conjointe optique/traitement
pour un imageur compact à capacité 3D**

JURY

Président :	François Goudail	IOGS
Rapporteurs :	Olivier Haeberlé	MIPS
	Philippe Réfrégier	Institut Fresnel
Examineurs :	Véronique Serfaty	DGA
	Éric Thiébaud	Observatoire de Lyon
	Jérôme Idier	IRCCyN
	Frédéric Champagnat	ONERA
Invité :	Guy Le Besnerais	ONERA

Directeur de thèse : Jérôme IDIER
Laboratoire : IRCCyN
Co-encadrant : Frédéric CHAMPAGNAT
Laboratoire : ONERA

N°ED : 503-182

Remerciements

Je tiens tout d'abord à remercier l'ensemble des membres du jury pour l'intérêt qu'ils ont porté à mes travaux de thèse et pour les différents échanges que nous avons eus, qui m'ont permis d'avoir différents points de vue sur mes travaux. Je remercie en particulier Olivier Haeberlé et Philippe Réfrégier d'avoir accepté d'être les rapporteurs de mon manuscrit ainsi que François Goudail, tout d'abord pour avoir accepté de présider mon jury, mais également pour les nombreuses discussions enrichissantes que nous avons eues auparavant, lors des comités d'évaluation de thèse. Enfin, je le remercie également pour l'opportunité qu'il m'a donné de faire des enseignements, cette expérience m'a en effet beaucoup apporté, tant sur le plan scientifique que personnel.

La conception conjointe d'un imageur implique plusieurs domaines scientifiques et j'ai eu le plaisir de travailler avec de nombreuses personnes dans le cadre de mes travaux de thèse. Je remercie tout d'abord Jérôme Idier d'avoir accepté de diriger ma thèse. Malgré la contrainte géographique, sa grande réactivité m'a permis de profiter de son expérience et de ses conseils. Je remercie Frédéric Champagnat et Guy Le Besnerais, pour leurs compétences scientifiques, leur disponibilité et leur enthousiasme, qualités que j'ai pu apprécier tout au long de ma thèse. Je remercie également Guillaume Druart pour son implication dans mes travaux. Ses compétences et son dynamisme m'ont permis de finir ma thèse avec une validation expérimentale de CAM3D. Merci également à Aurélien Plyer, Martial Sanfourche, Valérie Leung, Sandra Vitorino et Anthelme Bernard-Brunel pour leurs contributions respectives, que ce soit pour les algorithmes qu'ils ont développés ou pour leur aide pour les acquisitions ou le calibrage des différents imageurs que j'ai pu utiliser. Je remercie également Roland Domel qui a conçu les éléments mécaniques de CAM3D, Marc Jacquart qui m'a aidé à faire des acquisitions et Martin Piponnier qui m'a permis d'utiliser son matériel.

Mes remerciements s'adressent également aux membres de l'Institut d'Optique avec qui j'ai eu l'occasion de collaborer : Thierry Avignon, Fabienne Bernard, Lionel Jacubowicz et Jacques Sabater. Grâce à eux, j'ai pu disposer d'images réelles produites par deux prototypes différents. J'espère pouvoir leur rendre l'aide qu'ils m'ont apportée lors de nouvelles collaborations. Par ailleurs, je remercie également Bertrand Le Saux, Laurent Mugnier, Nicolas Guérineau et Jean Taboury pour les discussions intéressantes que nous avons eues au cours de ma thèse. Enfin, je remercie Hervé Lachaud et Pierre Breesé, pour leur aide quant à la rédaction du brevet déposé au cours de ma thèse.

Je tiens ensuite à remercier l'ensemble des membres du DTIM pour le soutien que j'ai reçu au quotidien et pour la bonne ambiance dans laquelle j'ai pu travailler. Je remercie en particulier Gilles Foulon et Patrick Secchi pour m'avoir accueillie au sein de l'unité EVF du DTIM. Je remercie également Françoise et Fabrice, qui nous communiquent chaque jour leur bonne humeur, ingrédient essentiel à l'aboutissement d'une thèse ! Je remercie également mes trois collègues de bureau successifs : Guillaume, Caroline et Aurélien, vous m'avez chacun beaucoup apporté. Enfin, je tiens à mentionner les nombreux doctorants rencontrés durant ma thèse : Anne-Marie, Antoine, Evangeline, Riadh, Joseph, Christophe, Walid, Laure, Paul, Jeff, Thibaut et Hicham. Merci à ceux qui m'ont précédée pour leurs conseils et bon courage aux futurs docteurs !

Ces trois années de thèse m'ont permis d'apprécier la qualité et la convivialité du travail au sein du DTIM et j'ai eu la chance de d'obtenir un poste d'ingénieur de recherche dans ce département. Je remercie donc chaleureusement toutes les personnes qui ont contribué à mon embauche.

Enfin, je remercie les membres de ma famille et mes amis pour leur soutien et Marius pour son affection et ses encouragements quotidiens pendant ces trois ans.

Table des matières

Table des matières	x
Liste des figures	xviii
Liste des tableaux	xx
Choix des notations	xxi
Introduction	1
1 Algorithme d'estimation locale de FEP mono-image	13
1.1 État de l'art	14
1.1.1 Modèle de formation d'image	14
1.1.2 Estimation multi-images	15
1.1.2.1 Élimination de la scène par filtrage inverse	15
1.1.2.2 Différentiation des images dans l'espace réel	16
1.1.2.3 Utilisation des outils probabilistes	17
1.1.2.4 Modélisation du flou de défocalisation par un phénomène de diffusion	17
1.1.2.5 Projection des données dans un sous-espace orthogonal	18
1.1.2.6 Bilan	18
1.1.3 Estimation mono-image	19
1.1.3.1 Utilisation des techniques de déconvolution aveugle	19
1.1.3.2 Modélisation des contours de la scène par une fonction de Heaviside	20
1.1.3.3 Modélisation du flou de défocalisation par un phénomène de diffusion	21
1.1.3.4 Approches marginales	21
1.1.3.5 Sélection d'une FEP dans une famille de FEP potentielles	22
1.1.4 Bilan	24
1.2 Algorithme d'estimation de FEP dans le cas mono-image	24
1.2.1 Modélisation des données	25
1.2.2 Modélisation de la scène	25

1.2.3	Vraisemblance marginale dans le cas d'un a priori propre	26
1.2.4	Vraisemblance marginale dans le cas d'un a priori impropre	26
1.2.5	Dérivation d'un critère d'estimation de FEP	27
1.2.6	Description de l'algorithme GLDFD	28
1.2.7	Exemple illustratif 1D	29
1.2.8	Implémentation	29
1.2.9	Bilan	31
1.3	Validation de l'algorithme GLDFD sur des données simulées	31
1.3.1	Scènes respectant l'a priori gaussien	32
1.3.1.1	Estimation de la FEP	32
1.3.1.2	Estimation du paramètre α	32
1.3.1.3	Variations du dénominateur du critère GL pour différentes FEP	33
1.3.1.4	Variation du nombre de données	33
1.3.1.5	Comparaison avec le critère GCV	34
1.3.1.6	Absence de la vraie FEP dans la famille de FEP potentielles	35
1.3.2	Scènes naturelles	36
1.3.2.1	Performance d'estimation de profondeur sur des scènes naturelles	36
1.3.2.2	Comparaison des décompositions en Fourier et GSVD	36
1.3.3	Bilan	39
1.4	Comparaison avec plusieurs méthodes de la littérature	39
1.4.1	Flou de bougé	39
1.4.1.1	Images simulées	40
1.4.1.2	Images réelles	41
1.4.1.3	Conclusion	42
1.4.2	Estimation du flou de défocalisation dans le cas d'une optique classique	42
1.4.2.1	Images simulées	43
1.4.2.2	Images réelles	44
1.4.2.3	Conclusion	44
1.4.3	Flou de défocalisation dans le cas d'une optique à pupille codée	45
1.4.3.1	Images simulées	46
1.4.3.2	Images réelles	47
1.4.3.3	Conclusion	48
1.4.4	Bilan des comparaisons menées	48
1.5	Discussions	48
1.5.1	Temps de calcul	48
1.5.2	Hypothèse de distribution isotrope	49
1.5.3	Modélisation du schéma Bayer	49
1.6	Conclusion	50
2	Solutions optiques favorisant l'estimation de profondeur	53
2.1	Présentation des deux solutions optiques	54
2.1.1	Utilisation d'une pupille codée pour favoriser l'estimation de profondeur	54
2.1.1.1	État de l'art	54
2.1.1.2	Limitations théoriques	56
2.1.2	Utilisation d'une optique chromatique	58
2.1.2.1	Principe et intérêt	58
2.1.2.2	État de l'art	59

2.2	Protocole expérimental	59
2.2.1	Calibrage des FEP	59
2.2.1.1	Acquisition de l'image d'un point source	60
2.2.1.2	Méthode de Levin [Levin <i>et al.</i> , 2007b]	60
2.2.1.3	Méthode de Delbracio [Delbracio <i>et al.</i> , 2011]	60
2.2.2	Acquisition de cibles à une profondeur connue	61
2.2.3	Traitement des résultats	61
2.2.4	Sources d'erreur	62
2.2.4.1	Précision de la vérité terrain	62
2.2.4.2	Variation de la FEP dans le champ	63
2.2.4.3	Angle entre le plan du détecteur et le plan de la cible	63
2.2.4.4	Calibrage de la FEP	63
2.2.4.5	Bilan	66
2.3	Étude d'un imageur à pupille codée	66
2.3.1	Réalisation d'un prototype	66
2.3.2	Calibrage des FEP	66
2.3.3	Mesures de distance sur l'axe	69
2.3.4	Cartes de profondeur	72
2.3.4.1	Comparaison de cartes de profondeur avec un système actif	72
2.3.4.2	Scènes extérieures	72
2.3.5	Bilan	74
2.4	Imageur chromatique	74
2.4.1	Généralisation de l'algorithme mono-image au cas multi-image	75
2.4.2	Généralisation de l'algorithme GLDFD au cas MA-DFD	75
2.4.2.1	Critère de sélection de FEP	75
2.4.2.2	Validation sur des images simulées	76
2.4.3	Généralisation de l'algorithme GLDFD au cas SAC-DFD	77
2.4.3.1	Modélisation de la scène	78
2.4.3.2	Critère de sélection de FEP	81
2.4.3.3	Validation sur des images simulées	82
2.4.4	Réalisation expérimentale	83
2.4.4.1	Conception d'un imageur chromatique	83
2.4.4.2	Calibrage des FEP sur l'axe	84
2.4.4.3	Calibrage des FEP hors axe	85
2.4.5	Mesures de distance sur l'axe	86
2.4.6	Cartes de profondeur	89
2.4.7	Discussion	90
2.4.7.1	Variation de la FEP suivant le spectre de la scène	90
2.4.7.2	Restauration de l'image RVB	91
2.5	Conclusion	93
3	Modélisation de la précision d'estimation de profondeur	95
3.1	État de l'art	96
3.1.1	Variation de la FEP avec la profondeur	96
3.1.2	Influence des traitements	98
3.1.3	Outil statistique : la borne de Cramér Rao	98
3.2	Cas général	99
3.2.1	Rappel des modèles utilisés dans le chapitre 1	100
3.2.2	Information de Fisher lorsque la matrice de précision est régulière	100

3.2.3	Information de Fisher lorsque la matrice de précision est singulière .	101
3.2.4	Calcul numérique de la BCR	102
3.3	Étude de la performance d'estimation de profondeur dans le cas SA-DFD .	102
3.3.1	Calcul de la BCR dans le cas SA-DFD	103
3.3.1.1	Calcul direct	103
3.3.1.2	Calcul de la matrice de covariance par préfiltrage des données	103
3.3.1.3	Calcul en utilisant la décomposition en base de Fourier . .	105
3.3.1.4	Choix d'une méthode de calcul	106
3.3.2	Étude de la BCR dans le cas d'un imageur conventionnel	107
3.3.2.1	Analyse d'un exemple de courbe de σ_{BCR}	107
3.3.2.2	Analyse de la variation de σ_{BCR} avec les paramètres algorith- miques	109
3.3.2.3	Interprétation de la variation de σ_{BCR} avec les paramètres optiques	110
3.3.2.4	Approximation de σ_{BCR} par une formule analytique	112
3.3.2.5	Bilan	115
3.3.3	Validations empiriques à l'aide de l'algorithme GLDFD	116
3.3.3.1	Variation de la mise au point	116
3.3.3.2	Variation de l'ouverture	118
3.3.3.3	Bilan	121
3.3.4	Comparaison des performances d'un imageur à pupille codée et à pupille circulaire	122
3.3.4.1	Comportement théorique	122
3.3.4.2	Étude sur des données simulées	123
3.3.4.3	Études sur des données réelles	124
3.3.5	Bilan	124
3.4	Étude de la performance d'estimation de profondeur dans le cas MA-DFD .	124
3.4.1	Calcul de la BCR dans le cas MA-DFD	125
3.4.1.1	Calcul direct	125
3.4.1.2	Calcul de la matrice de covariance par préfiltrage des données	126
3.4.1.3	Calcul en utilisant la décomposition de Fourier	126
3.4.1.4	Exemple	127
3.4.2	Optimisation du rapport des flous entre les deux images	128
3.4.3	Études sur des images simulées	129
3.4.4	Bilan	130
3.5	Étude de la performance d'estimation de profondeur dans le cas SAC-DFD	130
3.5.1	Calcul de la BCR dans le cas SAC-DFD	130
3.5.1.1	Calcul direct	130
3.5.1.2	Calcul de la matrice de covariance par préfiltrage des données	131
3.5.1.3	Calcul en utilisant une décomposition en Fourier	132
3.5.1.4	Exemple	133
3.5.2	Influence du modèle de scène chromatique sur la BCR	134
3.5.3	Comparaison de l'écart-type empirique avec σ_{BCR} sur des images si- mulées	135
3.5.3.1	Comportement théorique	136
3.5.3.2	Étude sur des données simulées	136
3.5.3.3	Étude sur des données réelles	137
3.5.4	Bilan	138
3.6	Études complémentaires	139

3.6.1	Variation de la BCR avec les autres paramètres de la matrice de précision	139
3.6.2	Modélisation du recul de la scène	140
3.6.3	Effet du filtre de Bayer	142
3.7	Comparaisons théoriques d’imageurs estimant la profondeur par SA-DFD ou SAC-DFD	144
3.7.1	Comparaison de deux solutions optiques pour le même détecteur et le même algorithme d’estimation de profondeur	144
3.7.2	Variation de la performance des traitements pour un même imageur	144
3.8	Conclusion	146
4	Co-conception d’un imageur à capacité 3D	147
4.1	Introduction	148
4.1.1	Problématique	148
4.1.2	Optimisation de la précision de l’imageur conçu à l’IOGS	149
4.1.3	Paramètres optiques choisis pour le dimensionnement préliminaire	151
4.2	Co-conception de l’imageur CAM3D	152
4.2.1	Cahier des charges	152
4.2.2	Définition de critères de conception conjoint	152
4.2.3	Détermination des paramètres principaux de l’imageur	153
4.2.3.1	Choix du détecteur	153
4.2.3.2	Distance focale	154
4.2.3.3	Ouverture	155
4.2.3.4	Taille de la fenêtre de données traitée par l’algorithme	155
4.2.3.5	Bornes sur la position des plans de mise au point RVB	155
4.2.4	Dimensionnement préliminaire des plans de mise au point RVB	156
4.2.4.1	Liste de triplets de mise au points des canaux RVB potentiels	156
4.2.4.2	Modélisation de l’imageur	156
4.2.4.3	Comparaison exhaustive des triplets de mise au point potentiels	157
4.2.4.4	Dimensionnement préliminaire pour le détecteur Dalsa	158
4.2.4.5	Dimensionnement préliminaire pour le détecteur Stingray	160
4.2.4.6	Conclusion	161
4.2.5	Conception d’une première architecture optique	161
4.2.5.1	Choix d’une architecture optique	161
4.2.5.2	Définition d’une fonction de mérite	161
4.2.5.3	Optimisation d’une première architecture	162
4.2.6	Optimisation conjointe	162
4.2.6.1	Récupération des FEP simulées par Zemax	163
4.2.6.2	Optimisation manuelle de l’objectif chromatique	163
4.3	Caractéristiques de l’objectif chromatique co-conçu	164
4.3.1	Paramètres des lentilles	164
4.3.2	Position des plans de mise au point RVB	164
4.3.3	Qualité de l’image hors axe	165
4.3.3.1	Variation de la FTM et de la FEP dans le champ	165
4.3.3.2	Aberrations chromatique longitudinale et latérale	167
4.3.4	Performance d’estimation de profondeur sur l’axe	167
4.3.5	Réalisation de l’imageur	167
4.3.5.1	Tolérancement et réalisation des lentilles	167

4.3.5.2	Mécanique	168
4.4	Caractérisation de la précision expérimentale	168
4.4.1	Calibrage des FEP sur l'axe	169
4.4.2	Précision sur l'axe avec le détecteur Dalsa	170
4.4.3	Précision sur l'axe avec le détecteur Stingray	170
4.4.4	Bilan	171
4.4.5	Cartes de profondeur	172
4.5	Discussion	173
4.5.1	Introduction des critères de performance dans la fonction de mérite de Zemax	173
4.5.2	Optimisation de la forme de la pupille codée	173
4.5.3	Variation de la FEP dans le champ	174
4.5.4	Limitation du repliement	175
4.5.5	Un pas de calibrage plus fin	175
4.6	Conclusion	176
	Conclusion générale	179
	Annexes	185
A	Dérivation de la vraisemblance dans le cas d'un a priori impropre	187
A.1	Définition d'un VIG	187
A.2	Propriétés des VIG par transformation linéaire	188
A.3	Définition d'une vraisemblance associée à un VIG	189
A.4	Cas d'un VIG filtré et bruité	190
B	Simulation de FEP à partir des formules d'optique de Fourier	193
C	Génération de scènes vérifiant l'a priori gaussien	195
C.1	Principe général	195
C.2	Pratique	195
D	Implémentation du critère GL à partir de la GSVD	197
D.1	Principe	197
D.2	Exemple simple	198
D.3	Expression du critère GL à partir de la GSVD	199
E	Approximation analytique de la BCR dans le cas mono-image	201
F	Scènes naturelles	205
	Bibliographie	215
	Publications et communications	217

Liste des figures

1	Modélisation de l'imageur.	xxii
2	(a) Caméra TOF SwissRanger 4000 développée par MESA Imaging. (b) Kinect diffusée par Microsoft pour la console de jeu XBOX 360.	2
3	(a) Exemple de système stéréoscopique (b) Caméra plénoptique développée par Lytro.	3
4	Le flou d'un objet dépend de sa position par rapport au plan de mise au point.	4
5	Illustration du lien entre la position d'un objet ponctuel et la largeur du flou de défocalisation.	4
6	Démarche de co-conception et organisation du manuscrit.	9
7	(a) Pupille codée proposée dans [Levin <i>et al.</i> , 2007b]. (b) Principe de l'utilisation d'une optique chromatique en DFD : la FEP de défocalisation est différente dans chaque bande spectrale.	10
1.1	(a) Modélisation de la scène par une fonction de Heaviside. (b) Principe de la méthode de [Elder et Zucker, 1998; Bae et Durand, 2007]. (c) Principe de la méthode de [Zhuo et Sim, 2009; Zhuo et Sim, 2011]	20
1.2	Schéma bloc de l'algorithme d'estimation de profondeur GLDFD.	29
1.3	Exemple de courbes du critère GL en fonction de α pour trois FEP dans le cas d'une scène constituée d'une marche ((a) et (b)) ou d'une scène uniforme ((c) et (d)).	30
1.4	Résultats d'estimation de profondeur avec le critère GL dans le cas de scènes vérifiant l'a priori gaussien et pour différentes valeurs de α pouvant s'interpréter comme l'inverse du rapport signal à bruit.	32
1.5	Résultats de l'estimation de α par l'algorithme GLDFD.	33
1.6	Variation du dénominateur du critère GL en fonction de α	33
1.7	Écart absolu moyen en % à la valeur vraie en fonction de la largeur des fenêtres traitées par l'algorithme GLDFD.	34
1.8	Comparaison des performances d'estimation de profondeur des critères GCV et GL sur les mêmes données dans le cas de scènes vérifiant l'a priori gaussien.	35
1.9	Résultats d'estimation de profondeur, en utilisant la famille de FEP ayant servi à simuler les images (famille F) ou la famille décalée (famille F').	35
1.10	Quatre scènes naturelles utilisées pour tester l'algorithme GLDFD.	36

1.11	Résultats de l'estimation de profondeur avec l'algorithme GLDFD sur des extraits des scènes (a) à (d) présentées à la figure 1.10.	37
1.12	Résultats d'estimation de profondeur à l'aide du critère GL implémenté à l'aide de la décomposition en Fourier et de la décomposition GSVD pour des fenêtres de tailles 21×21	37
1.13	Résultats d'estimation de profondeur à l'aide du critère GL implémenté à l'aide de la décomposition en Fourier pour différentes tailles de fenêtres et de la décomposition GSVD pour des fenêtres de tailles 21×21 , en enlevant de la famille de FEP potentielles les FEP les plus étroites (profondeurs inférieures à 2.4 m).	38
1.14	Résultat d'estimation de profondeur avec le critère GL et la décomposition de Fourier sur les scènes (a) à (d) de la figure 1.10.	38
1.15	Résultats d'estimation de FEP obtenus avec le critère GL et la méthode de [Chakrabarti <i>et al.</i> , 2010] moyennés à partir d'images simulées pour différentes valeurs de bruit.	40
1.16	(a) Image réelle associée à la référence [Chakrabarti <i>et al.</i> , 2010]. Résultats d'estimation de FEP avec : (b) la méthode [Chakrabarti <i>et al.</i> , 2010], (c) l'algorithme GLDFD. Les régions en noir correspondent aux régions non texturées rejetées par le filtre de Canny.	41
1.17	(a) Image réelle associée à la référence [Chakrabarti <i>et al.</i> , 2010]. (b) Résultats d'estimation de FEP avec : (b) de la méthode [Chakrabarti <i>et al.</i> , 2010], (c) l'algorithme GLDFD. Les régions en noir correspondent aux régions non texturées rejetées par le filtre de Canny.	41
1.18	Résultat d'estimation du flou de bougé obtenu avec l'algorithme GLDFD sur une image réelle avec deux objets mobiles dans des directions différentes. . .	42
1.19	Deux scènes utilisées dans la comparaison entre la méthode [Zhuo et Sim, 2011] et l'algorithme GLDFD.	43
1.20	Résultats d'estimation de FEP obtenus en utilisant la méthode [Zhuo et Sim, 2011] et l'algorithme GLDFD sur les deux scènes présentées à la figure 1.19.	44
1.21	(a) Deux images réelles issues de la base d'images de [Zhuo et Sim, 2011]. Résultats obtenu en utilisant : (b) la méthode [Zhuo et Sim, 2011], (c) le critère GL. Les régions en blanc correspondent aux régions homogènes rejetées par le filtre de Canny.	45
1.22	Exemples de FEP simulées en utilisant la pupille codée de [Levin <i>et al.</i> , 2007b].	46
1.23	Comparaison des résultats d'estimation de FEP obtenus avec la méthode de Martinello et l'algorithme GLDFD sans bruit sur des scènes naturelles.(a) Vérité. (b) et (c) Résultats de [Martinello et Favaro, 2011b] avec respectivement $n_\lambda = 200$ et $n_\lambda = 400$. (d) Résultats de l'algorithme GLDFD.	47
1.24	Résultats d'estimation de FEP obtenus avec la méthode de Martinello avec le critère GL avec un bruit d'écart-type 0.01 sur des scènes naturelles.(a) Vérité. (b) et (c) Résultats de [Martinello et Favaro, 2011b] avec respectivement $n_\lambda = 200$ et $n_\lambda = 400$. (d) Résultats du critère GL.	47
1.25	(a) Images issues de [Levin <i>et al.</i> , 2007b]. (b) Résultats de l'estimation de profondeur par la méthode de [Levin <i>et al.</i> , 2007b] avant l'étape de régularisation. (c) Résultats obtenus avec l'algorithme GLDFD. Les régions en noir correspondent aux régions homogènes rejetées par le filtre de Canny.	48
1.26	Exemple de schéma Bayer situé sur la matrice d'un détecteur couleur. Illustration issue de http://it.wikipedia.org/wiki/Schema_Bayer	49

2.1	Exemples de pupilles proposées dans la littérature :(a) [Gottesman et Feni- more, 1989],(b) et (c)[Hiura et Matsuyama, 1998], (d) et (e)[Zhou et Nayar, 2009a], (f)[Veeraraghavan <i>et al.</i> , 2007] , (g) et (h)[Zhou et Nayar, 2009b] pour deux niveaux de bruit différents, (i)[Levin <i>et al.</i> , 2007b], (j)[Morrison, 2011].	56
2.2	Taille géométrique $ \epsilon $ de la FEP en pixel pour un imageur de distance focale 20 mm, de nombre d'ouverture 3 et de pixels de $7\ \mu\text{m}$, pour une mise au point à 2 m. (a) Illustration de l'ambiguïté sur l'estimation de profondeur. (b) Illustration de la profondeur de la région de l'espace où la profondeur ne peut être estimée précisément. La région en rouge correspond à la zone où la variation de la FEP n'est plus observable.	57
2.3	Illustration d'une erreur d'estimation de profondeur liée à l'ambiguïté du flou de défocalisation de part et d'autre du plan de mise au point. L'échelle de couleur correspond à la largeur de l'écart-type des gaussiennes constituant la famille de FEP.	57
2.4	Exemple de variation du diamètre géométrique du flou pour les trois canaux d'une optique chromatique.	58
2.5	(a) Mire utilisée dans la méthode [Levin <i>et al.</i> , 2007b]. (b) Mire utilisée dans [Delbracio <i>et al.</i> , 2011].	61
2.6	Exemples de cibles utilisées pour évaluer l'estimation de profondeur.	62
2.7	Estimation de profondeur sur des images simulées avec une famille de FEP potentielles constituées respectivement : (a) des FEP théoriques, (b) des FEP théoriques bruitées afin de simuler l'erreur de calibrage.	64
2.8	Histogrammes des résultats d'estimation de profondeur obtenus pour trois pas de calibrages. (a) Les scènes sont générées suivant l'a priori gaussien et la profondeur vraie est à 2.7 m.(b) Les scènes sont générées suivant l'a priori gaussien et la profondeur vraie est à 2.2 m. (c) La scène est naturelle et la profondeur vraie est à 2.2 m.	65
2.9	Imageur à pupille codée.	67
2.10	Calibrage des FEP par : (a) un point source, (b) la méthode de [Levin <i>et al.</i> , 2007b] et (c) la méthode de [Delbracio <i>et al.</i> , 2011]. De haut en bas les profondeurs sont 2.7, 3.2 et 3.5 m.	67
2.11	Comparaison des FTM en intensité des pupilles calibrées suivant les diffé- rentes méthodes de calibrage pour une profondeur de 3.5 m.	68
2.12	Variation des FEP dans le champ pour trois profondeurs (de haut en bas 2 m, 3 m et 4 m). (a) sur l'axe, (b) $(\theta_h, \theta_v) = (0^\circ, 2^\circ)$, (c) $(\theta_h, \theta_v) = (0^\circ, 10^\circ)$, (d) $(\theta_h, \theta_v) = (10^\circ, 7^\circ)$. Le coin en haut à droite de l'image correspond aux positions angulaires $(\theta_h, \theta_v) = (20^\circ, 13^\circ)$	69
2.13	Résultats d'estimation de profondeur pour quatre scènes, en utilisant l'ima- geur à pupille codée. (a) Cible « Gauss », (b) Cible « Cercles », (c) Cible « Parterre », (d) Cible « Carreaux ».	70
2.14	(a) Performances d'estimation de profondeur pour l'imageur : (a) muni d'une pupille codée. (b) muni d'une pupille classique.	71
2.15	Performance d'estimation de profondeur (a) en négligeant l'effet du schéma Bayer. (b) En prenant en compte l'effet du schéma Bayer.	72
2.16	(a) Images acquises avec l'appareil photographique Nikon. Carte de profon- deur obtenue :(b) avec la Kinect et (c) en utilisant l'algorithme GLDFD. Les régions en noir correspondent aux régions homogènes rejetées en pré-traitement.	72
2.17	(a) Acquisitions en extérieur avec l'appareil à pupille codée. (b) Cartes de profondeurs obtenues avec l'algorithme GLDFD.	73

2.18	(a) Acquisition en extérieur avec l'imageur à pupille codée. Un câble est tendu entre les deux arbres. (b) Carte de profondeur obtenue l'algorithme GLDFD. Le câble apparaît nettement dans la carte de profondeur.	73
2.19	Validation de l'algorithme MA-GLDFD sur des scènes vérifiant l'a priori gaussien, pour différentes valeurs du paramètre α , paramètre correspondant à l'inverse du rapport signal à bruit.	76
2.20	Scènes utilisées pour valider l'algorithme MA-GLDFD sur des scènes naturelles.	77
2.21	Validation de l'algorithme MA-GLDFD sur les scènes naturelles (a) à (d) présentées à la figure 2.20.	77
2.22	Trois scènes naturelles	78
2.23	Répartition des pixels R/V R/B et B/V pour les images présentées à la figure 2.22.	79
2.24	Répartition des pixels L/C_1 L/C_2 et C_1/C_2 pour les images présentées à la figure 2.22.	80
2.25	Estimation de profondeur avec le critère GL-SAC dans le cas de scènes vérifiant l'a priori gaussien, pour différentes valeurs de α	82
2.26	Résultats de l'estimation de la profondeur avec l'algorithme SAC-GLDFD. Les scènes RVB sont des scènes naturelles.	83
2.27	Test de l'estimation de la profondeur. Les scènes sont naturelles et converties en niveau de gris	84
2.28	(a) Plan des lentilles constituant l'objectif chromatique. (b) Objectif chromatique combiné au détecteur Stingray.	84
2.29	FEP calibrées par la méthode de [Delbracio <i>et al.</i> , 2011] pour trois profondeurs : (a) 4.7 m, (b) 2.7 m et (c) 1.9 m.	85
2.30	Variation de la FEP pour différentes positions angulaires de la mire (θ_h, θ_v) par rapport à l'axe optique, pour trois profondeurs (de haut en bas 1 m, 2.7 m et 5 m. (a) sur l'axe (b) $(0^\circ, 1^\circ)$, (c) $(0^\circ, 3^\circ)$ et (d) $(0^\circ, 5^\circ)$ (e) $(4^\circ, 5^\circ)$. Le coin en haut à droite de l'image correspond aux angles $(10^\circ, 8^\circ)$	85
2.31	(a) Découpe de l'image en 9 régions où la FEP est supposée constante. (b) FEP calibrées dans les 9 régions de l'image pour les trois canaux à 3.9 m.	86
2.32	Résultats d'estimation de profondeur sur quatre cibles colorées. (a) Cible « Fleurs », (b) Cible « Parterre », (c) Cible « Carreaux », (d) Cible « Ville ».	87
2.33	Résultats d'estimation de profondeur pour les cibles (a) « Fleurs ». (b) Cible « Parterre ». (c) Cible « Carreaux » imprimées en noir et blanc.	88
2.34	Résultat d'estimation de profondeur en utilisant l'algorithme MA-GLDFD sur la cible « Fleurs » de la figure 2.6(b) imprimée en couleur.	88
2.35	(a) Acquisitions avec l'imageur chromatique. (b) Cartes de profondeur obtenues avec la Kinect. (c) Cartes de profondeur brutes obtenues avec l'algorithme SAC-GLDFD.	89
2.36	(a) Simulation de FEP polychromatiques associées aux spectres complet des canaux RVB. (b) Simulation de FEP monochromatique respectivement à 0.64μ , 0.55μ m et 0.48μ m.	90
2.37	Variation des coefficients (a) $a_{p,B}$, (b) $a_{p,V}$ et (c) $a_{p,V}$ avec la profondeur.	92
2.38	(a) Image RVB non corrigée. (b) Image RVB après transfert de fréquence suivant l'équation (2.35).	92
3.1	Comparaison des valeurs de σ_{BCR} obtenues par les trois méthodes de calcul : le calcul direct, le calcul par préfiltrage des données et le calcul par la décomposition en Fourier, dans le cas de FEP gaussiennes.	107

3.2	(a) Valeurs théoriques de σ_{BCR} . (b) Variation de $ \tau $, l'écart-type normalisé des gaussiennes modélisant les FEP, en fonction de la profondeur.	108
3.3	Variation de σ_{BCR} en fonction de : (a) la largeur du support contenant la FEP, (b) le nombre de données N , (c) le pas de dérivation δ et (d) l'inverse du rapport signal à bruit α	109
3.4	(a) Variation de σ_{BCR} à 3 m pour un imageur dont tous les paramètres sont fixés à l'exception de la distance focale (la position du détecteur est adaptée pour avoir un plan de mise au point fixe). (b) Variation de σ_{BCR} à 3 m pour un imageur dont tous les paramètres sont fixés à l'exception du rayon de la pupille. (c) (resp.(d)) Variation de σ_{BCR} (resp. de $ \tau $) avec la profondeur pour plusieurs valeurs de distances focales, les autres paramètres optiques étant fixés sauf d_{det} qui est adapté pour avoir une mise au point à 2 m. (e) (resp.(f)) Variation de σ_{BCR} (resp. de $ \tau $) avec la profondeur pour plusieurs valeurs de rayon de la pupille, les autres paramètres optiques étant fixés. . .	111
3.5	(a) Variation de σ_{BCR} pour un même imageur avec des plans de mise au point différents. (b) Variation de τ pour les deux mises au point.	112
3.6	(a) $\Delta_{\tilde{d}}$ et $\Delta_{\tilde{h}}$ calculés par FFT (b) Résultats des approximations de $\Delta_{\tilde{d}}$ et $\Delta_{\tilde{h}}$ par les équations (3.57) et (3.58). (c) Différence en valeur absolue entre l'approximation et le calcul par FFT.	113
3.7	Comparaison des valeurs de σ_{BCR} obtenues par la formule (3.67) et par le calcul par décomposition en Fourier.	115
3.8	Comparaison des valeurs de σ_{BCR} avec l'écart-type empirique σ_e des résultats d'estimation de profondeur obtenus avec l'algorithme GLDFD : (a) sur des scènes vérifiant l'a priori gaussien, (b) sur des scènes naturelles.	117
3.9	(a) Valeurs de σ_{BCR} obtenues avec les caractéristiques du tableau 3.3 pour une mise au point à 1.5 m et une mise au point à 1.8 m. (b) Écart-type empirique σ_e obtenu sur la cible « Gauss ». (c) Biais empirique.	119
3.10	Écart-type obtenu avec l'algorithme GLDFD pour différentes valeurs d'ouverture à 3 m. (a) Les scènes sont générées suivant l'a priori gaussien et le niveau de bruit est tel que $\alpha = 0.001$. (b) Les scènes sont générées à partir d'extraits de scènes naturelles avec un écart-type de bruit égale à 0.01. Les valeurs d'écart-type sont comparées aux valeurs de σ_{BCR}	120
3.11	(a) σ_{BCR} calculé à partir de FEP simulées et des caractéristiques optiques de l'objectif Nikon présenté dans le tableau 3.3.(b) Écart-type empirique σ_e pour les deux ouvertures. (c) Biais empirique.	121
3.12	Comparaison des valeurs de σ_{BCR} pour un imageur à pupille codée et un imageur conventionnel de mêmes caractéristiques.	122
3.13	Comparaison de σ_{BCR} et de l'écart-type obtenu avec l'algorithme GLDFD dans le cas d'un imageur à pupille codée (pc) et à pupille classique (pnc) : (a) sur des scènes respectant l'a priori gaussien et $\alpha = 0.001$, (b) sur des scènes naturelles avec $\sigma_b = 0.01$	123
3.14	Comparaison des performances d'un objectif avec et sans pupille codée.(a) σ_{BCR} calculé en prenant la moyenne des α estimés sur les fenêtres à une profondeur donnée. (b) Écart type empirique. (c) Biais empirique.	125
3.15	(a) Comparaison des courbes de σ_{BCR} par les trois méthodes de calcul : le calcul direct, le préfiltrage des données et la transformée de Fourier. (b) Écart-type τ en pixel des gaussiennes simulant la FEP pour chaque profondeur et les deux mises au point.	127

3.16	(a) Variation de σ_{BCR} en fonction du rapport des flous de défocalisation entre deux images σ_2/σ_1 . (b) Courbes de $\sigma_{\text{BCR}}^2 = \text{CRB}$ obtenues dans [Rajagopalan et Chaudhuri, 1998].	128
3.17	Comparaison des valeurs de σ_{BCR} avec les valeurs d'écart-type empirique obtenues : (a) sur des données simulées à partir de scènes vérifiant l'a priori gaussien. (b) sur des données simulées à partir de scènes naturelles.	129
3.18	(a) Comparaison des trois méthodes de calcul de σ_{BCR} pour des FEP gaussiennes : calcul direct, préfiltrage des données et décomposition en base de Fourier. (b) Variations de l'écart-type de la gaussienne représentant la FEP de chaque canal RVB en fonction de la profondeur.	134
3.19	(a) Variation de σ_{BCR} dans le cas SAC-DFD pour trois valeurs de μ_c . (b) Ecart entre les valeurs de σ_{BCR} calculées pour trois images sans considérer une corrélation entre les scènes en utilisant la formule de l'équation (3.115) et la formule de σ_{BCR} pour $\mu_c = 1$. (c) comparaison des valeurs de σ_{BCR} calculées dans le cas SAC-DFDs avec $\mu_c = 1$ et $\mu_c \rightarrow 0$ avec respectivement le cas de scènes considérées comme décorréelées et le cas MA-DFD.	135
3.20	(a) Valeurs de σ_{BCR} en fonction de la profondeur pour le système optique chromatique dont les paramètres sont décrits dans le tableau 3.6. (b) Variation de $ \epsilon $ avec la profondeur pour les trois plans de mise au point correspondant aux canaux RVB.	136
3.21	Comparaison des valeurs de σ_{BCR} et des valeurs d'écart-type empirique obtenues avec l'algorithme SAC-GLDFD pour des images simulées : (a) à partir de scènes vérifiant l'a priori gaussien, (b) à partir de scènes naturelles. . . .	137
3.22	(a) Variation du diamètre géométrique de la FEP $ \epsilon $ en pixel avec la profondeur pour les trois canaux RVB. (b) Valeurs de σ_{BCR} . (c) Résultats d'estimation de profondeur expérimentaux sur la cible « Parterre ».	138
3.23	Comparaison des valeurs de σ_{BCR} en tenant compte uniquement de l'estimation de p et en tenant compte de l'estimation de α et σ_b^2 , pour deux valeurs de α : (a) $\alpha = 0.001$ et (b) $\alpha = 0.01$	140
3.24	(a) Signal aléatoire représentant la scène à 1 m. (b) Variation de l'écart-type des gradients avec la profondeur.	141
3.25	Effet de la modélisation du recul de la scène sur les courbes de σ_{BCR} dans les cas : (a) SA-DFD, (b) MA-DFD et (c) SAC-DFD.	142
3.26	Effet de la prise en compte du schéma Bayer dans les trois cas (a) SA-DFD, (b) SAC-DFD et (c) MA-DFD.	143
3.27	Comparaison des performances théoriques d'un imageur chromatique et d'un imageur achromatique sur le même détecteur couleur tri-CCD. Dans les deux cas, l'estimation de profondeur est réalisée à l'aide de l'algorithme SAC-GLDFD.	145
3.28	Comparaison de l'estimation de profondeur en utilisant une optique chromatique et chaque canal séparément ou en combinant l'information des trois canaux. Le détecteur est un détecteur tri-CCD.	145
4.1	Principe de l'imageur à capacité 3D recherché. Les illustrations proviennent du prototype d'imageur chromatique présenté dans le chapitre 2.	148

4.2	Comparaison des performances théoriques et expérimentales de l'imageur chromatique conçu à l'IOGS, pour deux positions du détecteur différentes. Les figures (a) (c) et (e) correspondent aux résultats théoriques et expérimentaux obtenus avec la position du détecteur utilisée dans le chapitre 2. Les figures (b) (d) et (f) présentent les mêmes résultats mais pour la position optimisée du détecteur. Les droites verticales en pointillés correspondent à la position des plans de mise au point des canaux RVB.	151
4.3	Comparaison des allures des FEP simulées : (a) par les formules d'optique physique (FEP_{OP}), (b) par une fonction gaussienne (FEP_g). Le plan de mise au point est situé à 1.6 m, $\rho=0.31$ et les FEP sont simulées pour une profondeur de 3 m avec les paramètres du détecteur Dalsa et les paramètres de l'imageur chromatique fixés à la section 4.2.3. (c) Coupe horizontale des FEP pour les deux modélisations.	157
4.4	Illustration des performances du système minimisant le critère C1. (a) Variation de la taille géométrique des FEP pour les trois canaux RVB en fonction de la profondeur. (b) Variation du critère σ_{BCR} en fonction de la profondeur.	158
4.5	Illustration des performances du système maximisant le critère C2. (a) Variation de la taille géométrique des FEP pour les trois canaux RVB en fonction de la profondeur. (b) Variation du critère σ_{BCR} en fonction de la profondeur.	159
4.6	Illustration des performances du système correspondant à un compromis entre les critères C1 et C2. (a) Variation de la taille géométrique des FEP pour les trois canaux RVB en fonction de la profondeur. (b) Variation du critère σ_{BCR} en fonction de la profondeur.	159
4.7	Spectres des filtres RVB présents sur le détecteur Dalsa, d'après les données du constructeur.	162
4.8	Plan des lentilles de l'architecture choisie pour CAM3D sous Zemax. Les surfaces sont numérotées de 1 à 10. La surface (5) correspond à la pupille et la surface (10) au plan du détecteur. Nous indiquons pour chaque lentille la nature du verre.	163
4.9	Caractéristiques des lentilles de l'objectif chromatique co-conçu fixées dans Zemax.	165
4.10	(a) Diamètre contenant 80% de l'énergie encerclée de la FEP (b) Valeur de la FTM à la fréquence de Nyquist du détecteur Dalsa, soit 33 mm^{-1} , en fonction de la distance pour les trois canaux RVB.	165
4.11	(a) FTM du canal rouge à 3.7 m pour différentes valeurs du champ. (b) FTM du canal vert à 2.7 m. (c) FTM du canal bleu à 2.2 m. Le trait noir correspond à la fréquence de Nyquist du détecteur Dalsa, située à 33 mm^{-1}	166
4.12	(a) FEP à $(0^\circ, 0^\circ)$. (b) FEP à $(0^\circ, 5^\circ)$. (c) FEP à $(0^\circ, 10^\circ)$. (d) FEP à $(7^\circ, 7^\circ)$. De haut en bas les profondeurs sont respectivement 1.2 m, 1.8 m et 4.4 m. Le pas d'échantillonnage est de $7.4 \mu\text{m}$	166
4.13	(a) Variation de la distance focale Δf en fonction de la longueur d'onde, par rapport la référence du canal vert à $0.55 \mu\text{m}$. (b) Variation de la position du point image Δy en fonction de l'angle de champ du point source.	167
4.14	Performance théorique d'estimation de profondeur de l'objectif co-conçu (a) avec le détecteur Dalsa (b) avec le détecteur Stingray.	168
4.15	Monture conçue au DOTA pour l'objectif chromatique.	168
4.16	Photographie de l'objectif chromatique co-conçu combiné :(a) au détecteur Dalsa (b) au détecteur Stingray.	169

4.17	Comparaison des FEP simulées à partir du logiciel Zemax et des FEP calibrées par la méthode de [Delbracio <i>et al.</i> , 2011], pour un point source situé à 4 m. (a) et (b) FEP simulées et calibrées avec le détecteur Dalsa. (c) et (d) FEP simulées et calibrées avec le détecteur Stingray. Les FEP sont celles produites sur chaque canal R,V et B.	169
4.18	Résultats d'estimation de profondeur obtenus sur l'axe avec l'objectif chromatique co-conçu et le détecteur Dalsa pour les cibles (a) « Fleurs », (b) « Ville », (c) « Parterre », (d) « Carreaux ».	170
4.19	Résultats d'estimation de profondeur obtenus sur l'axe avec le détecteur Stingray pour les cibles (a) « Fleurs », (b) « Gauss » (c) « Parterre » (d) « Carreaux ».	171
4.20	Exemples de cartes de profondeur obtenues avec l'imageur CAM3D. L'échelle des profondeurs est en m. (a) Image RVB. (b) Carte de profondeur produite par la Kinect. (c) Carte de profondeur obtenues avec l'algorithme SAC-GLDFD.	172
4.21	Performances théoriques d'estimation de profondeur de l'imageur CAM3D pour différentes formes de pupilles.	174
4.22	(a) Pupille annulaire. (b) Pupille codée.	174
4.23	FEP des canaux RVB à 8 m pour différentes formes de pupilles. (a) Pupille circulaire. (b) Pupille annulaire. (c) Pupille codée.	175
4.24	Résultats de différentes approches de régularisation sur des images produites par un imageur à pupille codée. (a) Image RVB. (b) Carte de profondeur produit par la Kinect. (c) Carte de profondeur brute obtenue avec l'algorithme GLDFD. (d) Filtrage médian. (e) <i>Graphcut</i> (f) Fusion.	182
B.1	Modélisation de la FEP en optique de Fourier.	193
C.1	Génération d'une fenêtre de scène correspondant à l'a priori gaussien isotrope.	196
E.1	Coupes des fonctions $\alpha\Delta_{\tilde{d}}$, $\Delta_{\tilde{h}}$, $\alpha\Delta_{\tilde{d}} + \Delta_{\tilde{h}}$ et f suivant l'axe $\nu = 0$, pour $\tau = 2$ et $\alpha = 0.001$	202
E.2	Comparaison de la formule analytique approchée et du calcul numérique de B et de σ_{BCR} . (a) Comparaison des valeurs numériques de B calculées à partir de l'approximation (E.17) où par la formule (E.3). (b) Comparaisons des valeurs numériques de σ_{BCR} calculées à partir de l'équation (3.59) ou à partir de l'équation approchée (E.18).	204
F.1	Scènes naturelles utilisées pour simuler des images réelles.	205

Liste des tableaux

1.1	Paramètres de l'optique et du détecteur utilisés pour simuler les FEP d'un imageur idéal.	31
1.2	Caractéristiques du système optique idéal utilisé dans les simulations d'un imageur à pupille codée.	46
2.1	Moyenne et écart-type tronqués associés aux histogrammes de la figure 2.8.	64
2.2	Caractéristiques du système optique idéal utilisé dans les simulations. . . .	82
2.3	Écart moyen à la vérité et écart-type obtenus sur un ensemble d'images simulées pour différentes valeur de μ_c	83
3.1	Paramètres utilisés pour simuler les FEP d'un imageur et calculer les valeurs de σ_{BCR} présentées à la figure 3.2.	108
3.2	Paramètres utilisés pour simuler les FEP d'un imageur idéal dans le cas d'une modification de la mise au point.	116
3.3	Paramètres de l'appareil photographique Nikon utilisés pour calculer σ_{BCR} lors d'une modification de la mise au point.	118
3.4	Paramètres utilisés pour tracer la figure 3.12	122
3.5	Paramètres de l'imageur chromatique simulé.	133
3.6	Paramètres de l'imageur chromatique simulé.	136
4.1	Caractéristiques principales des deux détecteurs couleurs Stingray et Dalsa.	154
4.2	Triplets de mise au point maximisant la précision d'estimation de profondeur (critère C1), maximisant la capacité à produire une image de grande profondeur de champ (critère C2) et triplet présentant un compromis entre les deux critères.	158
4.3	Triplets correspondant à un compromis entre les critères C1 et C2 pour différentes valeurs du paramètre α	160
4.4	Plans de mise au point optimaux suivant les critères C1, C2 et un compromis entre les deux critères pour le détecteur Stingray.	160
4.5	Triplets optimisant les critères C1 et C2 pour différentes valeurs du paramètre AXCL.	164
4.6	Performances d'estimation de profondeur moyennes sur la plage de profondeurs comprises entre 1 et 5 m pour les quatre cibles de la figure 4.19. Le biais moyen est calculé en moyennant la valeur absolue des biais expérimentaux.	172

E.1 Paramètres utilisés pour tracer la figure E.2.	204
--	-----

Notations

Nous présentons ci-dessous les abréviations utilisées dans ce manuscrit ainsi que les notations associées aux différentes grandeurs que nous allons considérer.

ABRÉVIATIONS

FEP : Fonction d'Étalement du Point, correspond à l'image d'un point source.

DFD : Depth from Defocus, technique d'estimation de profondeur à partir d'une estimation du flou de défocalisation.

SA-DFD : Single Acquisition Depth From Defocus, technique de DFD avec une seule acquisition.

MA-DFD : Multiple Acquisitions Depth From Defocus, technique de DFD avec plusieurs acquisitions de la même scène.

SAC-DFD : Single Acquisition Chromatic Depth From Defocus, technique de DFD avec une acquisition issue d'un imageur chromatique.

PARAMÈTRES DU SYSTÈME OPTIQUE ET DU DÉTECTEUR

$F/\#$: nombre d'ouverture.

f : distance focale de l'optique.

d_{det} : distance entre l'optique et le détecteur.

p_0 : distance entre le plan de mise au point et l'optique.

p : distance entre l'objet et l'optique.

R : rayon de la pupille de l'optique.

p_x : taille du pixel (considéré carré).

ϵ : diamètre géométrique du flou de défocalisation.

τ : écart-type de la gaussienne modélisant le flou de défocalisation, normalisé en pixel.

Ces notations sont en partie définies à la figure 1.

PARAMÈTRES CARACTÉRISANT UNE ACQUISITION DU DÉTECTEUR

(m, n) : indexation des pixels de l'image.

\mathbf{y} : vecteurs contenant les pixels d'une fenêtre de l'image dans la représentation lexicogra-

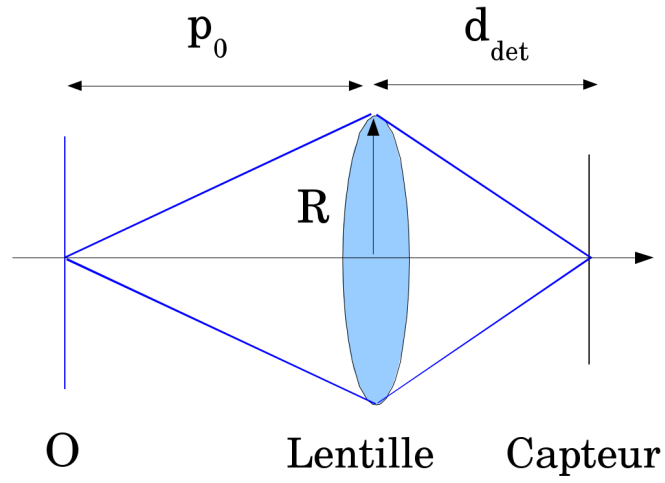


Fig. 1: Modélisation de l'imageur.

phique.

N : dimension du vecteur \mathbf{y} .

\mathbf{x} : vecteurs contenant les pixels d'une fenêtre de la scène dans la représentation lexicographique.

M : dimension du vecteur \mathbf{x} .

h : matrice 2D contenant la FEP échantillonnée sur le détecteur.

H : matrice de convolution associée à la FEP.

\mathbf{b} : vecteur aléatoire modélisant le bruit d'acquisition.

Introduction

Les travaux présentés dans ce manuscrit ont été effectués à l'Onera au sein du Département de Traitement de l'Information et Modélisation, en collaboration avec le Département d'Optique Théorique et Appliquée de l'Onera ainsi qu'avec l'Institut de Recherche en Communications et Cybernétique de Nantes. Ces travaux, financés par la DGA et l'Onera, sont rattachés au projet de recherche fédérateur SPIDER, dirigé par G. Le Besnerais, dont l'objectif est d'ajouter de nouvelles capacités de perception à des appareils autonomes tels que les micro-drones, grâce au développement d'imageurs innovants combinant un système optique et des algorithmes de traitement d'images conçus conjointement. Cette nouvelle démarche de conception, désignée par le terme de co-conception ou conception conjointe, consiste à optimiser simultanément le système optique et les traitements appliqués à l'image, afin de concevoir des imageurs plus compacts et plus performants que les imageurs traditionnels. L'enjeu est alors de définir de nouveaux critères de conception combinant les paramètres de l'optique, du détecteur et des algorithmes de traitement d'images et caractérisant la performance globale du système en lien avec la fonctionnalité ajoutée à l'imageur.

Dans ce contexte, l'objectif de cette thèse est de concevoir conjointement un imageur capable de produire une image couleur (RGB) ainsi qu'une carte de profondeur de la scène. Ce nouveau type d'imageur à capacité 3D est désigné par le terme de caméra RGB-D, la lettre D faisant référence à l'anglais *Depth* qui désigne la profondeur. Une des motivations de ce travail est le développement croissant des caméras RGB-D depuis quelques années, que ce soit dans le domaine militaire ou civil. En effet, dans le domaine de la défense, l'essor de la robotique (drones, robots terrestres et marins) a rendu nécessaire le développement de techniques d'extraction d'une information 3D pour la localisation et la reconstruction de l'environnement de ces appareils au cours de leur mission. De plus, les nouvelles fonctionnalités associées à des technologies grand public, tels que le téléphone portable intelligent ou les nouvelles consoles de jeu interactives, impliquent de plus en plus souvent l'environnement 3D autour de l'utilisateur. Enfin, un des motifs de la co-conception d'un capteur à capacité 3D est que cette capacité peut être évaluée expérimentalement de manière simple, en utilisant par exemple un télémètre comme référence de profondeur.

Dans cette introduction, nous présentons brièvement les différentes voies technologiques et algorithmiques possibles pour concevoir une caméra RGB-D. Parmi ces voies, nous avons choisi d'étudier une solution utilisant un imageur passif monovoie, dans laquelle la profondeur est estimée par une identification locale du flou de défocalisation. Nous montrons en

effet que cette approche permet d'envisager des systèmes très compacts et très économes, et qu'elle ouvre aussi des opportunités intéressantes de co-conception. Nous faisons ensuite une revue des travaux relevant de la co-conception en distinguant les imageurs conçus en combinant des algorithmes de traitement d'images et une optique conventionnelle, approche désignée par le terme de *computational photography*, et les imageurs conçus en combinant des algorithmes de traitement d'images à une optique innovante, démarche désignée alors par le terme de *computational imaging*. Cette revue rapide nous permet d'identifier les outils nécessaires à la co-conception d'un imageur à capacité 3D qui ne sont pas présents dans l'état de l'art et qui ont constitué nos axes de travail. Nous présentons ensuite nos principales contributions et le plan de ce manuscrit.

Caméra RGB-D

L'objectif de cette section est de présenter les avantages et les inconvénients des différents procédés optiques et algorithmiques permettant l'estimation de la profondeur.

Caméra RGB-D active

Parmi les différents procédés d'estimation de profondeur, les procédés dits « actifs » exploitent un signal, le plus souvent infrarouge, envoyé sur la scène. Deux exemples assez répandus sont les caméras *time of flight* (TOF) et la Kinect distribuée par Microsoft qui sont représentées toutes les deux à la figure 2. La caméra TOF repose sur le principe de télémétrie et ne produit qu'une image en niveau de gris de la scène. La Kinect, quant à elle, utilise une illumination de la scène par une mire en infrarouge et la déformation de cette mire avec la profondeur permet d'estimer la profondeur en temps réel. La Kinect est sans doute à ce jour l'exemple de caméra RGB-D active le plus connu du grand public.

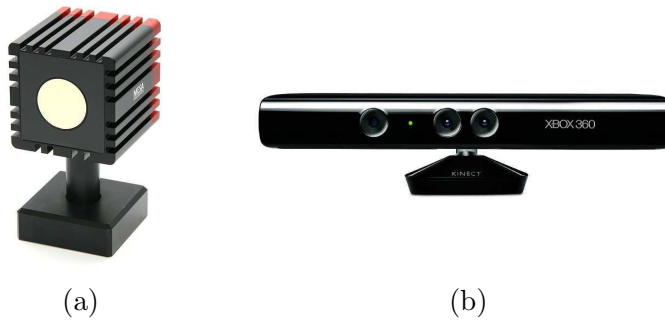


Fig. 2: (a) Caméra TOF SwissRanger 4000 développée par MESA Imaging. (b) Kinect diffusée par Microsoft pour la console de jeu XBOX 360.

Les procédés actifs ont l'avantage de pouvoir estimer la profondeur quelle que soit la texture de la scène, avec une grande précision. Par contre, les caméras actives possèdent un émetteur en plus de l'imageur ce qui augmente l'encombrement et la consommation en énergie. De plus, ces caméras ne sont pas furtives car le signal émis est détectable. Enfin, des interférences possibles entre les signaux empêchent l'utilisation de ces caméras sur plusieurs appareils voisins, ou en extérieur à cause de l'illumination infrarouge du Soleil. Cette famille de caméra RGB-D n'a donc pas été retenue, les applications aux drones étant un axe privilégié du projet SPIDER dans lequel s'inscrit cette thèse.

Caméra RGB-D passive

Contrairement aux procédés actifs, les procédés dits « passifs » estiment la profondeur uniquement à partir d’une ou de plusieurs images de la scène.

Au sein de cette famille, les systèmes multi-voies reposent sur l’effet de parallaxe entre deux ou plusieurs images de la même scène. Les premiers systèmes reposant sur cet effet sont les systèmes stéréoscopiques, composés de deux imageurs identiques, décalés comme dans l’exemple présenté à la figure 3(a). La profondeur est obtenue par une mesure locale de disparité entre les deux images et des cartes de profondeur de très grande qualité ont été obtenues grâce à ces systèmes [Hirschmuller et Scharstein, 2009]. La précision d’estimation de distance est alors proportionnelle à la distance entre les optiques, ce qui limite la compacité de l’imageur. De plus, l’estimation de distance dépend du calibrage du positionnement relatif des deux optiques, positionnement qui peut varier, par exemple à cause des vibrations liées au fonctionnement du porteur. Enfin, l’acquisition des deux images doit se faire de manière simultanée, ce qui nécessite une synchronisation précise des deux imageurs.

Très récemment, de nouveaux systèmes optiques multi-voies permettant d’estimer la profondeur grâce à l’effet de parallaxe ont été développés. Il s’agit des caméras plénoptiques qui combinent un système optique à une seule lentille avec une matrice de microlentilles placée au niveau du détecteur [Adelson et Wang, 1992; Ng *et al.*, 2005; Georgiev, 2010]. Le premier système optique forme une image intermédiaire qui est ensuite imagée par les matrices de microlentilles. Ce type de caméra apparaît aujourd’hui sur le marché des imageurs grand public avec par exemple le produit de la start-up Lytro¹, présenté à la figure 3(b). Ces appareils présentent l’intérêt de pouvoir modifier le plan de mise au point des images après l’acquisition. De plus, la distance entre le plan focal intermédiaire et les matrices de microlentilles permet d’exalter la parallaxe et de donner une information 3D sur la scène [Bishop et Favaro, 2011]. Cependant, outre leur complexité technologique, les caméras plénoptiques sont très gourmandes en pixels et la résolution de l’image se trouve affectée.

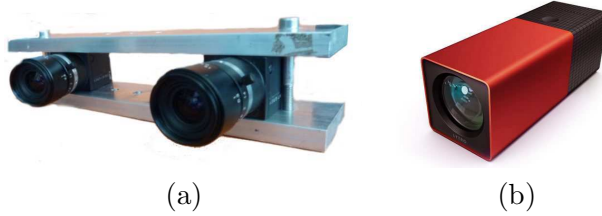


Fig. 3: (a) Exemple de système stéréoscopique (b) Caméra plénoptique développée par Lytro.

Enfin, il existe des systèmes passifs estimant la profondeur en n’utilisant qu’une seule voie optique. Ces systèmes, dits mono-voies, utilisent l’information de profondeur contenue dans le flou de défocalisation. En effet, le flou d’un objet dans une image dépend de sa position vis-à-vis du plan de mise au point, comme l’illustre la figure 4. La figure 5 montre que la relation entre la largeur géométrique du flou et la position de l’objet peut être définie par des formules d’optique géométrique. L’image d’un objet ponctuel, appelée Fonction d’Étalement du Point (FEP) est quasi-ponctuelle lorsque cet objet est placé comme le point O_1 dans le plan de mise au point de l’imageur, et sa largeur dépend des effets de la diffraction. Par contre, lorsque cet objet est placé en dehors du plan de mise au point comme O_2 , la FEP correspond à un flou de défocalisation de diamètre géométrique ϵ qui dépend

¹<http://www.lytro.com/>.



Fig. 4: Le flou d'un objet dépend de sa position par rapport au plan de mise au point.

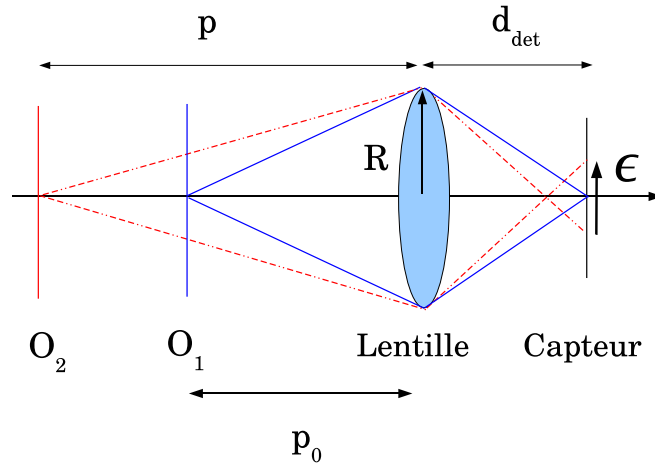


Fig. 5: Illustration du lien entre la position d'un objet ponctuel et la largeur du flou de défocalisation.

de la position du point source par rapport à l'imageur et des paramètres de l'optique. Cette FEP est responsable du flou de l'image d'un objet étendu. Le diamètre ϵ peut s'exprimer suivant la relation (1), dont les notations sont définies en figure 5 et avec f la distance focale de la lentille :

$$\epsilon = 2Rd_{det} \left(\frac{1}{f} - \frac{1}{p} - \frac{1}{d_{det}} \right). \quad (1)$$

Dans la littérature, il existe deux grandes familles de méthodes exploitant le flou de défocalisation pour estimer la profondeur : les techniques de *Depth From Focus* (DFF) et les techniques de *Depth From Defocus* (DFD).

Les techniques de DFF utilisent une série d'acquisitions réalisées en modifiant la position du plan de mise au point, puis visent à sélectionner parmi la série d'acquisitions, la mise au point pour laquelle l'objet est net. Ceci indique la valeur de d_{det} , pour laquelle le flou de défocalisation est de taille minimale ce qui permet alors d'estimer la valeur de p connaissant la distance focale de l'imageur f [Nair et Stewart, 1992; Nayar, 1992; Xiong et Shafer, 1993; Subbarao et Choi, 1995]. L'inconvénient de ces techniques est qu'elles impliquent une modification de la position du détecteur au cours des acquisitions et donc une consommation importante en énergie. De plus, cette méthode demande de stocker un grand

nombre d’images, de l’ordre d’une vingtaine. Enfin la scène doit être statique au cours de la série d’acquisitions.

Les techniques de DFD classiques reposent sur une estimation locale de la profondeur grâce à des algorithmes d’estimation locale du flou de défocalisation relatif entre deux acquisitions obtenues en modifiant les configurations optiques de l’imageur comme l’ouverture, la focale ou la mise au point [Pentland, 1987; Rajagopalan et Chaudhuri, 1997b; Subbarao, 1988; Subbarao et Surya, 1994]. Ces techniques demandent donc d’acquérir et de stocker moins d’images que les techniques de DFF mais nécessitent encore que la scène soit restée statique entre les deux acquisitions. Récemment des méthodes de DFD mono-image ont également été développées [Veeraraghavan *et al.*, 2007; Levin *et al.*, 2007b; Zhou et Nayar, 2009a; Martinello *et al.*, 2010; Zhuo et Sim, 2011; Martinello et Favaro, 2011b]. Dans ce cas, une seule image est nécessaire ce qui simplifie l’acquisition mais, comme nous le détaillerons dans le chapitre 1, en contrepartie les traitements d’estimation du flou sont plus complexes.

Après cette rapide revue des différents procédés optique et algorithmique existants permettant l’estimation de la profondeur, il s’avère que les méthodes mono-voies passives par DFD mono-image sont celles qui respectent le mieux des contraintes de faible consommation, de discrétion et de compacité. De plus, plusieurs équipes de chercheurs [Veeraraghavan *et al.*, 2007; Levin *et al.*, 2007b; Zhou et Nayar, 2009a; Martinello *et al.*, 2010; Martinello et Favaro, 2011b] ont montré que la forme de la pupille de l’optique, pouvait être optimisée afin de favoriser l’estimation de profondeur, ce qui ouvre des opportunités intéressantes pour la co-conception.

Co-conception

La combinaison d’une optique avec un détecteur numérique a permis de modifier l’image après son acquisition, ce qui a conduit au développement de nombreux algorithmes de traitement d’images qui accroissent les performances des imageurs. Pour cela, les algorithmes les plus anciens de traitement d’images utilisent les images en sortie d’une optique traditionnelle optimisée en vue d’avoir la meilleure résolution possible compte tenu de la taille des pixels. Or, récemment, plusieurs équipes de recherche provenant des communautés optique et traitement d’image ont montré qu’une modification astucieuse de l’optique associée aux traitements peut permettre d’améliorer encore les performances des imageurs.

Dans la communauté optique, le développement de nouveaux imageurs combinant une optique innovante et des traitements conduisent à l’apparition d’un nouveau thème de recherche intitulé COSI (*Computational optical sensing and imaging*)[Brady *et al.*, 2008]. Dans cette thématique, de nouveaux systèmes optiques sont conçus, pour des applications diverses telles que l’holographie, l’imagerie hyperspectrale, la microscopie ou la photographie.

Parallèlement, dans le domaine du traitement d’image, deux thèmes de recherche se sont développés : la *computational photography* et le *computational imaging*. En *computational photography*, le principe est de trouver de nouveaux procédés d’acquisition d’image, à partir d’optiques conventionnelles associées à des algorithmes innovants, ceci afin d’améliorer la qualité de l’image, ou d’extraire une information sur la scène. Puisque ces procédés font intervenir des optiques conventionnelles, l’intérêt de la *computational photography* est que les solutions proposées sont en général peu coûteuses. Dans le *computational imaging*, l’objectif est de trouver des solutions optiques innovantes conçues pour améliorer les performances des algorithmes de traitement d’images. L’image en sortie de l’optique est alors souvent exploitable sans les traitements. Dans les paragraphes suivants, nous présentons brièvement

plusieurs articles provenant des deux communautés optique et traitement d’images pour des applications en photographie, en distinguant les approches de type *computational photography* de celles relevant du *computational imaging*. Nous nous intéressons en particulier aux références qui traitent de la correction du flou de bougé, qui apparaît lorsqu’un objet de la scène ou l’appareil se déplacent durant le temps d’acquisition du détecteur, et celles qui traitent de la correction du flou de défocalisation, notamment dans le but d’étendre la profondeur de champ. De nombreuses autres références relevant de ces deux approches sont décrites dans les deux revues récentes [Nayar, 2011; Zhou et Nayar, 2011].

Computational photography

Dans ce domaine, une application largement développée est l’estimation et la correction du flou de bougé. Par exemple, [Yuan *et al.*, 2007; Sorel et Sroubek, 2009] proposent d’utiliser deux images : une image présentant du flou de bougé et une image nette, obtenue avec un temps d’exposition plus court et donc avec un bruit supérieur. Dans [Nayar et Ben-Ezra, 2004; Tai *et al.*, 2008] une caméra haute résolution avec un temps d’acquisition long, donc sensible au flou de bougé, est associée à une caméra basse résolution avec un temps d’acquisition court, qui estime le flou de bougé à l’aide d’une série d’acquisitions à des instants différents. Une deuxième caméra basse résolution est ajoutée dans [Li *et al.*, 2008] au système précédent afin d’estimer une carte de profondeur par effet stéréoscopique. Dans [Zhuo *et al.*, 2010], les informations hautes fréquences d’une image avec flash sont utilisées pour améliorer l’image acquise avec une illumination naturelle. Dans [Joshi *et al.*, 2010] un vélocimètre est associé à une caméra ce qui permet d’estimer précisément son déplacement lors de l’acquisition et donc de corriger le flou de bougé sur l’image. Enfin, en superrésolution multi-images, plusieurs images ayant un décalage subpixelique sont utilisées afin d’augmenter la résolution de l’image au delà de la limite définie par les pixels du détecteur [Champagnat *et al.*, 2009; Létienne, 2012].

Concernant le flou de défocalisation, l’ensemble des techniques de DFD et DFF peuvent être considérées comme des exemples de *computational photography* car une nouvelle information est extraite à partir d’images acquises avec un appareil conventionnel mais avec des configurations optiques différentes. Ces techniques sont détaillées dans ce manuscrit, en particulier dans le chapitre 1. Par ailleurs, dans la référence [Hasinoff et Kutulakos, 2008] une série d’acquisitions est effectuée avec différents temps d’expositions, ouvertures et mises au point, tout en maintenant l’illumination constante, afin de produire une image de grande profondeur de champ et une carte de profondeur, par la combinaison de ces images.

Computational imaging

Dans le cas du flou de bougé, les auteurs de [Levin *et al.*, 2008; Cho *et al.*, 2010] déplacent la caméra dans la direction du mouvement pendant le temps d’acquisition. En déplaçant la caméra suivant une trajectoire parabolique, ils montrent que le flou de bougé est alors identique pour tous les objets de la scène, quelle que soit leur vitesse de déplacement. Le flou devient alors uniforme dans toute l’image et celle-ci est restaurée par une unique déconvolution.

Par ailleurs, de nombreux articles utilisent une approche de type *computational imaging* afin d’augmenter la profondeur de champ d’un imageur. Une première approche consiste à rendre la FEP invariante avec la profondeur. Ainsi, tous les objets appartenant à une large plage de profondeur ont le même flou et un traitement unique de déconvolution permet d’obtenir une image de grande profondeur de champ. Par exemple dans [Cathey et Dowski, 2002; Diaz *et al.*, 2009] une lame de phase cubique est introduite dans l’imageur et

cette lame est optimisée afin d’augmenter la profondeur de champ de l’image restaurée. Par ailleurs, dans [Nagahara *et al.*, 2008] le détecteur est déplacé durant le temps d’acquisition afin d’obtenir un flou de défocalisation uniforme. Enfin, dans [Cossairt, 2010], une optique chromatique et un détecteur noir et blanc sont utilisés. Le chromatisme permet d’avoir une FEP polychromatique invariante avec la profondeur. Une seconde approche pour étendre la profondeur de champ de l’imageur vise à renforcer la fluctuation de la FEP avec la profondeur. Le traitement consiste alors à estimer localement la FEP, ce qui permet d’obtenir une carte de profondeur de la scène puis de corriger localement l’image du flou de défocalisation, ce qui augmente la profondeur de champ de l’image restaurée. Par exemple, l’introduction d’une lame de phase dans l’optique est proposée dans [Dowski Jr et Cathey, 1994; Johnson *et al.*, 2000]. Il est également possible de modifier la forme du diaphragme de l’optique afin de favoriser les traitements d’estimation de FEP en utilisant par exemple une pupille codée, comme dans les références [Veeraraghavan *et al.*, 2007; Levin *et al.*, 2007b; Zhou et Nayar, 2009a; Martinello *et al.*, 2010; Martinello et Favaro, 2011b]. Enfin dans [Guichard *et al.*, 2009] une optique chromatique est associée à un traitement de transfert de hautes fréquences inter-canal ce qui permet également d’augmenter la profondeur de champ de l’image.

Critères de conception conjoints

La conception conjointe de l’optique et des traitements nécessite de définir de nouveaux critères de conception qui prennent en considération à la fois les paramètres optique et traitement. Contrairement à la conception classique, ces critères ne caractérisent plus uniquement la qualité de l’image en sortie de l’optique. Parmi les travaux cités précédemment, qui proposent de nouveaux concepts d’imageurs, certains d’entre eux justifient le choix des paramètres optique et traitement par un critère de conception. Dans [Dowski Jr et Cathey, 1994] la borne de Cramér Rao est utilisée pour définir un critère que doivent respecter les paramètres de la lame de phase cubique pour favoriser la fluctuation de la FEP avec la profondeur. Dans la thèse [Diaz, 2011] les paramètres du masque de phase augmentant la profondeur de champ sont optimisés à partir d’un critère caractérisant le rapport signal à bruit de l’image après traitement. Ce critère est calculé à l’aide de l’erreur quadratique moyenne de l’image restaurée, moyennée sur une plage de profondeur. Dans [Stork et Robinson, 2008], la qualité de l’image après traitement est améliorée à l’aide d’un critère d’erreur de déconvolution, prenant en compte les paramètres de l’optique et des traitements. Ces paramètres sont optimisés à l’aide d’une boucle impliquant un logiciel de conception optique simulant les différentes solutions optiques. Dans [Levin *et al.*, 2007b], la forme de la pupille codée est optimisée en maximisant la distance de Kullback entre les pupilles à différentes échelles. Par ailleurs, dans [Zhou et Nayar, 2009b] les auteurs développent un critère d’optimisation de la forme d’une pupille codée qui minimise l’erreur de déconvolution afin de produire une image de grande profondeur de champ. Enfin, dans [Martinello et Favaro, 2011b] est décrit un critère de comparaison des performances d’estimation de profondeur de différentes formes de pupilles codées, calculé à partir d’une base de données apprise pour chaque forme de pupille.

Conception conjointe d’un imageur mono-voie passif à capacité 3D

Parmi l’état de l’art dans le domaine de la co-conception, il existe peu d’outils de conception conjointe d’un système mono-voie passif à capacité 3D. En effet, la construction de carte

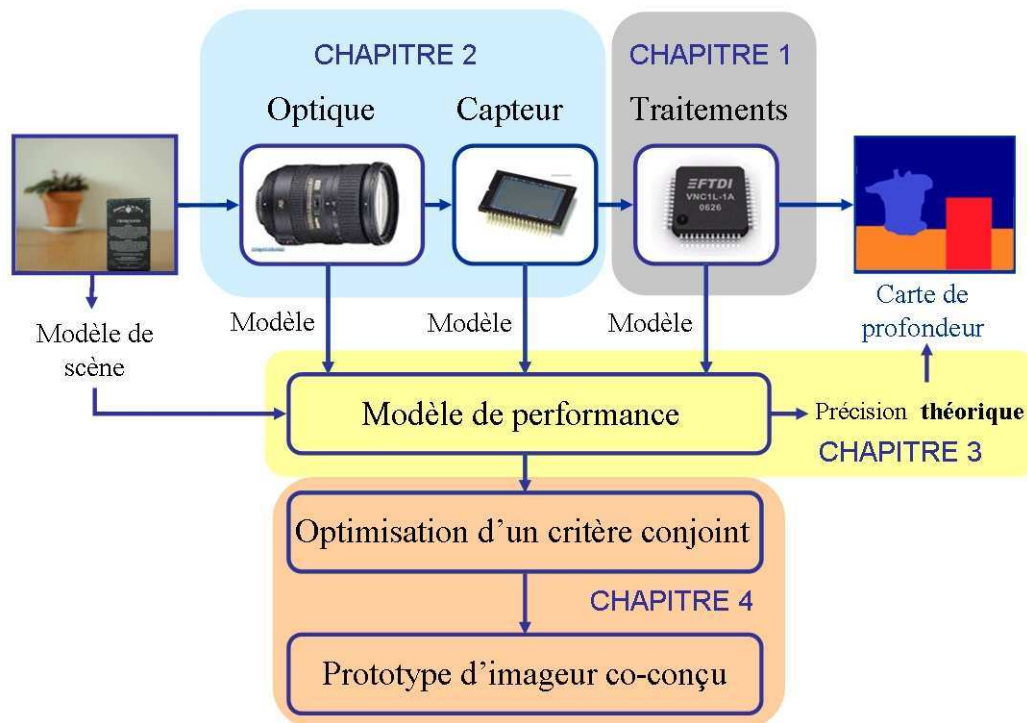
de profondeur n'est souvent qu'une étape intermédiaire des travaux dont l'objectif est l'augmentation de la profondeur de champ et, à l'exception de [Martinello et Favaro, 2011b], la précision d'estimation de profondeur n'est que rarement évaluée.

Le critère de conception ayant conduit à la pupille codée proposée dans [Levin *et al.*, 2007b] ne considère qu'une partie du système optique, à savoir la pupille, et non l'ensemble des paramètres de l'optique de l'imageur, ni des traitements. Par contre, dans [Martinello et Favaro, 2011b] un critère de comparaison de performance d'estimation de profondeur est proposé, construit à partir des FEP de l'imageur, ce qui permet de prendre en compte d'autres paramètres que la pupille. Cependant, le calcul de ce critère repose sur l'utilisation d'une base d'apprentissage, ce qui alourdit fortement les temps de calcul et le critère proposé n'est alors utilisé que pour comparer différentes pupilles. De plus, l'utilisation d'une base d'apprentissage réduit la généralité ainsi que la reproductibilité de la méthode, ce qui constitue un obstacle à la co-conception car le résultat de l'optimisation d'un imageur dépendra alors de la base choisie. Par ailleurs, les critères proposés dans [Levin *et al.*, 2007b; Martinello et Favaro, 2011b] sont des critères informationnels qui permettent de comparer des solutions optiques mais qui ne sont pas directement interprétés en termes de précision d'estimation de profondeur. Il manque donc à la littérature des outils de co-conception génériques permettant d'optimiser la performance d'estimation de profondeur d'un imageur monovoie passif, et c'est précisément un des objectifs de cette thèse.

Démarche et contributions

L'objectif principal de cette thèse est la conception d'un imageur monovoie passif à capacité 3D en utilisant un critère conjoint tenant compte des aspects de l'optique et des algorithmes de traitement d'images. Cette conception nécessite trois outils : un algorithme d'estimation de profondeur, une solution optique favorisant cette estimation et un modèle de performance. Ce modèle doit permettre de prévoir la précision d'estimation de profondeur théorique d'un imageur en fonction des paramètres de l'optique et des traitements. Ces trois outils ont été développés au cours de ces travaux de thèse puis ont été utilisés pour concevoir et réaliser un prototype d'imageur à capacité 3D. Le schéma de la figure 6 illustre notre démarche et fait le lien avec les différents chapitres de ce manuscrit.

Le chapitre 1 présente un nouvel algorithme d'estimation de la FEP locale à partir d'une seule image. Les algorithmes proposés dans la littérature reposent souvent sur une modélisation de la forme du flou de défocalisation par une gaussienne, ce qui limite les possibilités de modification de l'optique et donc les opportunités de co-conception. D'autres méthodes plus générales ont été proposées mais reposent sur un apprentissage à partir d'une base d'images, ce qui limite la reproductibilité et la généralité de l'algorithme d'estimation de profondeur. De plus, l'utilisation de ces techniques implique une dépendance de la précision vis-à-vis d'une base d'images particulière, ce qui n'est pas souhaitable lors de la phase de conception d'un imageur. Nous avons donc développé un algorithme non supervisé et suffisamment général pour être utilisé pour une large famille d'imageurs monovoie passifs. Cet algorithme, que nous désignons par GLDFD, pour *Generalized Likelihood Depth from Defocus*, a été conçu afin de pouvoir estimer la profondeur à partir d'une seule acquisition, cas que nous appellerons dans la suite SA-DFD pour *Single Acquisition Depth from Defocus*, qui est le cas le plus délicat du point de vue des traitements. L'algorithme GLDFD prend en entrée une famille de FEP potentielles produites par l'imageur et estime, sur une fenêtre de l'image, la FEP la plus probable. Chaque FEP étant associée à une profondeur, l'algorithme estime ainsi une carte de profondeur de la scène. La sélection de la FEP est réalisée à



l'aide de la minimisation d'un critère nommé GL, pour *Generalized Likelihood*, dérivé d'un calcul de maximum de vraisemblance. Ainsi, par rapport aux algorithmes d'estimation de profondeur de la littérature, l'algorithme que nous avons développé est plus générique, car il prend en entrée une famille quelconque de FEP potentielles. De plus, notre formalisme nous permet de généraliser cet algorithme au cas d'acquisitions multiples d'une scène unique, cas que nous désignerons dans la suite par MA-DFD, pour *Multiple Acquisitions Depth from Defocus*. Ces deux aspects laissent donc une grande liberté sur le choix de l'optique. Enfin, cet algorithme est facilement reproductible car il est non supervisé, ce qui est également un avantage pour la co-conception. Dans le chapitre 1, nous validons l'algorithme GLDFD sur des images simulées puis sur des images réelles issues de la littérature. Puis nous le comparons à trois algorithmes de l'état de l'art ce qui permet d'illustrer sa généricité et sa robustesse.

de profondeur obtenues avec l'imageur à pupille codée et avec la Kinect, qui ont montré que nos cartes de profondeurs, bien que plus bruitées, sont cohérentes avec une carte fournie par ce système actif.

Nous avons ensuite développé une seconde solution optique permettant de favoriser l'estimation de profondeur. Cette solution consiste à combiner une optique non corrigée de l'aberration chromatique, un détecteur couleur et un algorithme dérivé de l'algorithme GLDFD. Comme illustré à la figure 7(b), la lentille chromatique permet d'obtenir une FEP spectralement variable. Le détecteur couleur fait l'acquisition de l'image dans différentes bandes spectrales (le plus souvent les bandes spectrales correspondant aux couleurs rouge, verte et bleue), et permet d'obtenir en une seule acquisition plusieurs images ayant un flou différent. Ces images, appelées aussi canaux, peuvent être combinées pour produire une estimation de profondeur. Le principe de cette estimation est connu et peut être rapproché des travaux de MA-DFD. Cependant, à notre connaissance, il n'y a pas dans la littérature d'imageur effectuant une estimation de profondeur par DFD à partir d'une optique chromatique. Pour réaliser cet imageur nous avons généralisé l'algorithme GLDFD au cas de l'acquisition d'images chromatiques, cas que nous désignerons dans la suite par SAC-DFD pour *Single Acquisition Chromatic Depth from Defocus*. Cet algorithme, nommé SAC-GLDFD, utilise le même formalisme que l'algorithme GLDFD mais une modélisation différente de la scène, faisant intervenir la corrélation entre les canaux. Grâce à une deuxième collaboration avec l'IOGS, nous avons pu disposer d'une optique chromatique qui a permis de valider notre algorithme. Le prototype d'imageur chromatique obtenu a ensuite été évalué sur des cibles en comparaison avec des moyens télémétriques. Ceci constitue donc la première réalisation d'un imageur chromatique à capacité 3D. L'inconvénient de la solution chromatique est qu'elle dégrade la qualité image, du fait de la résolution variable suivant les bandes spectrales. Nous présentons à la fin de ce chapitre quelques résultats de restauration d'images RVB obtenus par une méthode de transfert de hautes fréquences, sur des acquisitions réalisées avec l'imageur chromatique.

La mise en œuvre expérimentale de ces deux solutions, un imageur à pupille codée et un imageur chromatique, a permis d'illustrer l'intérêt de la solution chromatique. Nos résultats expérimentaux montrent en effet qu'une optique chromatique permet d'estimer la profondeur sur une région plus large de l'espace et avec une meilleure précision qu'une optique munie d'une pupille codée.

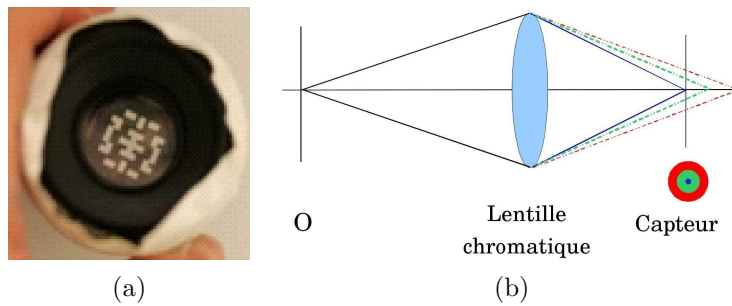


Fig. 7: (a) Pupille codée proposée dans [Levin *et al.*, 2007b]. (b) Principe de l'utilisation d'une optique chromatique en DFD : la FEP de défocalisation est différente dans chaque bande spectrale.

La co-conception d'un imageur à capacité 3D nécessite de pouvoir caractériser de manière théorique la performance d'estimation de profondeur d'un imageur, qu'il soit achromatique ou chromatique. C'est pourquoi, le chapitre 3 de ce manuscrit est consacré à un modèle théorique de performance, combinant les paramètres de l'optique et des traitements.

Ce modèle est basé sur un calcul original de la borne de Cramér Rao (BCR) dans les cas SA-DFD, MA-DFD et SAC-DFD. Il prend en entrée la ou les FEP associées à l'imageur pour une profondeur, le rapport signal à bruit et les paramètres de l'algorithme de traitement d'estimation de profondeur. Le modèle de performance renvoie alors la BCR, c'est-à-dire la précision théorique maximale de l'imageur à cette profondeur. À l'aide de ce modèle, le comportement de la précision d'estimation de profondeur est étudiée dans les cas SA-DFD, MA-DFD et SAC-DFD pour différentes valeurs des paramètres optiques tels que l'ouverture ou la mise au point. Les tendances prédites par les courbes théoriques sont ensuite vérifiées sur des données simulées puis expérimentalement, à l'aide de différents types d'imageurs, à savoir un appareil photographique classique, le prototype d'imageur à pupille codée et le prototype d'imageur chromatique. Ces vérifications ont montré que si la précision théorique n'est pas atteinte expérimentalement, les configurations préférentielles en termes de précision d'estimation de profondeur prévues par la théorie sont identiques à celles observées empiriquement.

Le chapitre 4 combine tous les acquis des chapitres précédents et présente la conception conjointe d'un imageur chromatique monovoie à capacité 3D. Cet imageur est optimisé grâce à notre modèle de performance, auquel nous avons adjoint une contrainte sur la qualité de l'image RVB après traitement, ceci afin d'obtenir un imageur produisant simultanément une carte de profondeur précise dans une gamme de profondeur définie préalablement et une image couleur de bonne qualité. Cette deuxième contrainte concerne précisément la capacité de produire une image de grande profondeur de champ, à l'aide par exemple d'un transfert de l'information haute fréquence contenue dans les différents canaux, elle peut donc être interprétée comme une profondeur de champ généralisée. L'optimisation de cet imageur a été réalisée en plusieurs étapes. Après avoir fixé les paramètres principaux de l'imageur tels que la focale et l'ouverture, un ordre de grandeur de la position optimale des plans de mise au point RVB a été évalué en utilisant un modèle simple de FEP gaussiennes. Puis à partir d'une première configuration de l'imageur chromatique conçue en collaboration avec le Département d'Optique Théorique et Appliqué de l'Onera (DOTA), le chromatisme et la mise au point de l'imageur ont été ré-optimisés en calculant les critères de conception à partir des FEP simulées dans le logiciel de conception optique Zemax. L'objectif chromatique ainsi conçu a ensuite été réalisé par le DOTA. Ensuite, nous avons calibré l'imageur et évalué expérimentalement sa performance d'estimation de profondeur sur deux détecteurs différents. Au delà de la performance de l'imageur obtenu, notamment pour l'un des détecteurs, la démarche de co-conception d'imageur monovoie passif à capacité 3D fait l'objet d'un dépôt de brevet.

L'ensemble des résultats de ces travaux de thèse sont repris dans la conclusion de ce manuscrit. Ainsi, à l'issue de ces travaux nous disposons d'un imageur à capacité 3D co-conçu reposant sur l'estimation du flou de défocalisation à partir d'un imageur chromatique. Nous présentons les améliorations possibles qu'il serait intéressant d'apporter aux traitements conduisant à la production d'une carte de profondeur et à une image RVB de bonne qualité. Il s'agit notamment de l'implémentation en GPU des algorithmes d'estimation de profondeur pour produire une carte de profondeur à une cadence proche de celle des images RVB. En effet, le calcul du critère de l'algorithme GLDFD, pour chaque FEP et pour chaque fenêtre de l'image, se prête bien à la parallélisation, ce qui permet d'accélérer les temps de calcul. Ce travail de parallélisation a fait l'objet d'un travail au sein du Département de Traitement de l'Information et Modélisation de l'Onera (DTIM), dont nous présentons les premiers résultats en termes de temps de calcul. D'autre part, les cartes de profondeurs

peuvent être régularisées notamment pour éliminer les valeurs de profondeur aberrantes. Plusieurs de ces techniques sont brièvement présentées et appliquées sur une carte de profondeur brute obtenue avec l'algorithme GLDFD. Enfin, il serait intéressant d'approfondir le traitement de restauration d'une image RVB, brièvement abordé dans le chapitre 2 pour que l'imageur chromatique que nous proposons puisse véritablement s'apparenter à une caméra RGB-D.

Nous revenons ensuite sur la démarche de co-conception utilisée et sur les études théoriques de concepts optiques qui pourraient compléter le travail réalisé dans cette thèse. Nous mentionnons enfin des exemples de nouvelles applications pour la méthode de conception conjointe.

Algorithme d'estimation locale de FEP mono-image

Introduction

L'estimation de profondeur à l'aide d'un imageur monovoie passif à capacité 3D repose sur la capacité d'estimer localement le flou de défocalisation, autrement dit la FEP, à partir d'une ou de plusieurs acquisitions. Pour être inclus dans un procédé de co-conception, l'algorithme d'estimation de profondeur doit pouvoir être utilisé avec une large famille de solutions optiques et ne doit donc pas dépendre d'une forme de FEP particulière. Dans la mesure du possible, il doit être non supervisé, car l'utilisation d'une base d'apprentissage crée un risque de sur-apprentissage et diminue la reproductibilité des résultats. Comme nous le verrons, les algorithmes d'estimation de FEP de la littérature ne possèdent pas ces deux caractéristiques. C'est pourquoi nous avons développé un nouvel algorithme d'estimation de FEP, utilisable pour une large famille d'optiques et non supervisé. Il est décrit ici dans le cas de l'estimation de FEP à partir d'une seule acquisition mais il est généralisé au cas multi-images dans le chapitre 2. L'algorithme que nous proposons repose sur une modélisation de la scène et de l'image dans un contexte bayésien et sur la définition d'un critère noté GL pour *Generalized Likelihood*, dérivé de l'estimation par maximum de vraisemblance, qui permet de sélectionner la FEP la plus probable parmi une famille de FEP potentielles. Ainsi, l'algorithme proposé, que nous nommons GLDFD, permet de traiter le cas de l'estimation du flou de défocalisation comme le cas du flou de bougé. Cet algorithme a fait l'objet d'une publication dans le congrès ICIP en 2011.

ORGANISATION DU CHAPITRE

Dans ce chapitre, nous commençons par présenter les méthodes d'estimation de FEP de l'état de l'art afin de positionner notre travail parmi les méthodes existantes. Puis nous décrivons les modèles de la scène et de l'image utilisés pour exprimer le critère GL, ainsi qu'une méthode d'implémentation de ce critère permettant le traitement d'une fenêtre de taille réduite sans approximation sur les bords. L'algorithme GLDFD est ensuite testé sur des images simulées, afin d'étudier ses performances en fonction de paramètres tels que le nombre de données, la variance du bruit ou la texture de la scène. Enfin l'algorithme est comparé à plusieurs algorithmes de la littérature.

1.1 État de l'art

Dans cette section nous présentons l'état de l'art concernant l'estimation de FEP. Une grande majorité des articles cités traitent de l'estimation du flou de défocalisation, qui est l'application principale de ce travail, mais certains d'entre eux concernent également l'estimation du flou de bougé. Ils sont présents dans cette étude bibliographique car ils ont contribué à l'élaboration de l'algorithme présenté dans ce chapitre.

1.1.1 Modèle de formation d'image

L'image y d'une scène x est classiquement modélisée par la relation :

$$y = h * x + b, \quad (1.1)$$

où h , appelée FEP, correspond à l'image d'un point source et b modélise le bruit d'acquisition. La FEP est la combinaison des contributions de l'optique, du détecteur ainsi que de la position et du déplacement de la scène. D'une manière très générale h est modélisée par :

$$h = h_{opt} * h_{det} * h_v. \quad (1.2)$$

où h_{opt} contient le phénomène de diffraction, les aberrations ainsi que la défocalisation, h_{det} est liée à l'échantillonnage de l'image sur le détecteur. Enfin h_v correspond au flou de bougé lié au déplacement de la scène ou de l'imageur pendant le temps d'acquisition. La relation (1.1) est valable localement, car la plupart des différentes sources de flou varient spatialement dans l'image.

Pour des déplacements dans les directions horizontales ou verticales, le flou de bougé peut être modélisé par une FEP à une dimension dans l'une de ces directions [Levin, 2007; Chakrabarti *et al.*, 2010]. Des déplacements plus complexes impliquent des FEP à deux dimensions. Par ailleurs, dans le cas où le flou de bougé est négligé, deux modélisations de la FEP sont couramment utilisées dans la littérature. La première consiste à modéliser la combinaison de h_{opt} et h_{det} par une fonction gaussienne. Ainsi h est couramment modélisée par la relation :

$$h(m, n) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(m^2+n^2)}{2\sigma^2}\right), \quad (1.3)$$

où le paramètre σ est défini par :

$$\sigma = \rho\epsilon = 2\rho R d_{det} \left(\frac{1}{f} - \frac{1}{p} - \frac{1}{d_{det}} \right), \quad (1.4)$$

où ρ un paramètre obtenu par calibrage et ϵ la taille géométrique du flou de défocalisation. Les paramètres p, f et d_{det} sont définis à la page [xxi](#). Notons que si la scène est placée devant le plan de mise au point, σ atteint des valeurs négatives, mais cela n'a pas d'influence sur la FEP car l'équation (1.4) fait intervenir σ^2 . Une autre modélisation de la FEP est également parfois utilisée. Elle consiste à modéliser h comme un disque dont les dimensions sont données par la relation :

$$h(m, n) = \begin{cases} \frac{4}{\pi\epsilon^2} & \text{si } m^2 + n^2 \leq (\epsilon/2)^2 \\ 0 & \text{si } m^2 + n^2 > (\epsilon/2)^2 \end{cases}. \quad (1.5)$$

Cette modélisation revient à négliger les effets de la diffraction et des aberrations. Ces deux modèles sont cependant insuffisants lorsque la pupille de l'imageur est modifiée pour favoriser l'estimation de profondeur [Levin *et al.*, 2007b]. Une modélisation plus fine consiste à utiliser les formules d'optique de Fourier, présentées dans l'annexe B. Cette modélisation permet d'utiliser une forme de pupille quelconque est de modéliser la FEP défocalisée en fonction des paramètres de l'optique et de la position du point source par rapport au plan de mise au point. Dans ce manuscrit nous utilisons ces formules pour un imageur dit « idéal », qui ne présente pas d'aberration, mais il est possible d'introduire des déphasages liées aux différentes aberrations de l'optique dans ce modèle.

Notre objectif est d'estimer localement la FEP. La difficulté est que l'image dépend de deux inconnues à savoir la FEP et la scène. Une première famille d'algorithmes utilise plusieurs images pour estimer h , mais des méthodes mono-images ont également été développées. Ces deux familles de méthodes sont présentées successivement dans cette section.

1.1.2 Estimation multi-images

Une première solution pour estimer la FEP de défocalisation consiste à utiliser plusieurs images produites par la même scène mais avec un flou différent, obtenues par un réglage différent de l'imageur. Nous désignons ce cas par le terme de MA-DFD pour *Multiple Images Depth from Defocus*. Cette solution permet alors de supprimer la scène du jeu d'inconnues. Dans les paragraphes suivants sont présentées les méthodes d'estimation de FEP de défocalisation dans le cas multi-images, qui sont regroupées suivant la nature de la technique d'estimation utilisée.

1.1.2.1 Élimination de la scène par filtrage inverse

Les premières méthodes d'estimation de flou de défocalisation reposent sur un filtrage inverse, à savoir sur la suppression des termes relatifs à la scène en calculant par exemple le rapport du spectre des deux images. En effet, si y_1 et y_2 sont deux acquisitions de la même scène x avec différentes configurations optiques, alors dans l'espace de Fourier ces grandeurs vérifient :

$$\tilde{y}_1(\mu, \nu) = \tilde{h}_1(\mu, \nu)\tilde{x}(\mu, \nu), \quad (1.6)$$

$$\tilde{y}_2(\mu, \nu) = \tilde{h}_2(\mu, \nu)\tilde{x}(\mu, \nu), \quad (1.7)$$

où h_1 est la FEP relative au premier mode d'acquisition et h_2 celle du second. De plus, en modélisant la FEP h à l'aide d'une gaussienne, suivant l'équation (1.3), la transformée de Fourier de la FEP s'écrit :

$$\tilde{h}(\mu, \nu) = \exp(-2\pi^2\sigma^2(\mu^2 + \nu^2)). \quad (1.8)$$

Le rapport des modules carrés des équations (1.6) et (1.7) implique la relation suivante :

$$4\pi^2(\sigma_2^2 - \sigma_1^2) = \frac{1}{(\mu^2 + \nu^2)} [\ln|\tilde{y}_1|^2(\mu, \nu) - \ln|\tilde{y}_2|^2(\mu, \nu)]. \quad (1.9)$$

Il est alors possible d'estimer la différence $\sigma_2^2 - \sigma_1^2$ à partir du spectre des deux images à une fréquence donnée. En réécrivant la relation (1.4) pour les deux modes d'acquisition et en combinant les deux équations obtenues pour éliminer le paramètre p , on peut montrer que les niveaux de flou absolu de chaque image sont reliés par la relation :

$$\sigma_1 = \gamma\sigma_2 + \eta, \quad (1.10)$$

où γ et η dépendent des paramètres optiques des deux modes d'acquisition.

La référence [Pentland, 1987] est l'un des premiers articles présentant une réalisation expérimentale de MA-DFD. Dans cet article, deux images acquises avec une ouverture différente sont utilisées, l'une a une ouverture ponctuelle, tandis que l'autre a une ouverture plus importante. Dans ce cas $\sigma_1 = 0$ et l'équation (1.9) renseigne directement sur la valeur du flou de la deuxième image. Dans [Subbarao, 1988], le concept de MA-DFD est généralisé au cas d'une variation finie ou infinitésimale d'un ou de plusieurs paramètres de l'optique tels que la focale, la mise au point ou l'ouverture. Avec la méthode de Subbarao, il n'est plus nécessaire d'avoir une image acquise avec une ouverture ponctuelle, qui introduit notamment une forte différence de flux entre les deux images, ainsi qu'un effet de diffraction plus important. Toutefois une modification de la distance focale ou de la mise au point nécessite des traitements sur les deux images, pour corriger notamment la différence de grandissement des différentes configurations de l'imageur. Subbarao rend également l'estimation du flou plus robuste en calculant l'intégrale :

$$\sigma_2^2 - \sigma_1^2 = \frac{1}{4\pi^2 A} \int_R \frac{1}{\mu^2 + \nu^2} [\ln|\tilde{y}_1|^2(\mu, \nu) - \ln|\tilde{y}_2|^2(\mu, \nu)] d\mu d\nu, \quad (1.11)$$

où R est une région du spectre où les densités spectrales de puissance des deux images sont différentes et A l'aire de cette région. La combinaison des équations (1.10) et (1.11) permet d'obtenir une équation du second degré en σ_2 en fonction de paramètres optiques fixés et des densités spectrales de puissances des deux images.

Dans [Bove Jr *et al.*, 1993], le phénomène de défocalisation est décrit en termes de variation d'entropie. Cette variation est calculée à partir de la transformée de Fourier des images :

$$H = \sum_{\mu=0}^{N-1} \sum_{\nu=0}^{N-1} |\ln|\tilde{y}_1(\mu, \nu)| - \ln|\tilde{y}_2(\mu, \nu)|| \quad (1.12)$$

Puis les valeurs de H mesurées sont comparées à une table de référence pour en déduire la taille du flou et donc la profondeur.

D'autres représentations fréquentielles sont utilisées dans la littérature pour estimer le flou relatif entre les deux images. Ainsi dans [Gokstorp, 1994] les images sont filtrées par des filtres de Gabor qui permettent d'estimer localement la fréquence, l'amplitude et la phase instantanée de l'image. La procédure d'estimation du flou est alors itérative et consiste à estimer à chaque étape le flou relatif entre les images, à convoluer l'image la plus nette par ce flou, puis à recalculer le flou relatif. Cette procédure est répétée de manière multi-échelle, jusqu'à ce que l'écart entre l'image la plus floue et l'image nette convoluée par le flou estimé soit nul. Dans [Rajagopalan et Chaudhuri, 1997b], le spectrogramme complexe ainsi que la pseudo distribution de Wigner sont proposés afin d'estimer le flou dans une représentation spatio-fréquentielle de l'image. Enfin, dans [Watanabe et Nayar, 1998] un ensemble de filtres permettant d'estimer la profondeur indépendamment de la texture sont proposés à partir d'une modélisation de la FEP par un disque.

1.1.2.2 Différentiation des images dans l'espace réel

En pratique, l'estimation de profondeur est réalisée sur des fenêtres des deux images, et le modèle des équations (1.6) et (1.7) est très approximatif car il ne prend pas en compte le sinuscardinal lié au fenêtrage. Une solution consiste à traiter les images dans l'espace réel. Dans [Hwang *et al.*, 1989], la profondeur est estimée à partir de deux images acquises avec

des positions du détecteur différentes. La relation de convolution entre la scène et l'image filtrée est dérivée par rapport à la position du détecteur d_{det} . En prenant le Laplacien du résultat, Hwang obtient :

$$\sigma \frac{\partial \sigma}{\partial d_{det}} = \frac{\frac{y_1(m,n) - y_2(m,n)}{d_{det1} - d_{det2}}}{0.5(\nabla^2 y_1(m,n) + \nabla^2 y_2(m,n))}. \quad (1.13)$$

Les formules d'optique géométrique telle que (1.4) exprimées pour les deux configurations permettent d'exprimer $\sigma \frac{\partial \sigma}{\partial d_{det}}$ en fonction de la profondeur. Remarquons que les images sont préfiltrées par une gaussienne pour limiter l'effet du bruit. Dans [Subbarao et Surya, 1993; Subbarao et Surya, 1994] la scène est décomposée en polynômes de degrés trois en fonction des variables spatiales m et n . Cette modélisation permet d'écrire le développement de Taylor de la scène et après plusieurs étapes de calcul, les auteurs définissent la différence entre les deux images acquises par la relation :

$$y_1(m,n) - y_2(m,n) = \frac{1}{4} (\chi_1^2 - \chi_2^2) \nabla^2 \left(\frac{y_1(m,n) + y_2(m,n)}{2} \right), \quad (1.14)$$

où χ est le moment d'ordre 2 de la FEP. Comme dans [Subbarao, 1988], une équation du second degré en χ_2 est obtenue en combinant l'équation (1.14) avec les formules d'optique géométrique faisant intervenir les paramètres optiques des deux configurations. La résolution de cette équation permet alors d'estimer la profondeur. Ici seul le moment d'ordre 2 de la FEP intervient sans impliquer une modélisation de celle-ci par une gaussienne.

1.1.2.3 Utilisation des outils probabilistes

Les techniques citées précédemment, reposant sur un filtrage inverse ou sur la différenciation des images, ont pour principal défaut la robustesse au bruit qui n'est pas pris en compte dans la modélisation de l'image. C'est pourquoi d'autres méthodes utilisant une modélisation du bruit ont été développées. Ces méthodes utilisent notamment des outils probabilistes tels que le maximum a posteriori (MAP) ou le maximum de vraisemblance (MV). Dans [Ens et Lawrence, 1993] un formalisme matriciel est utilisé pour estimer le flou relatif entre deux images et le problème d'estimation de FEP est formulé par la minimisation d'un critère quadratique régularisé en utilisant un a priori sur la forme de la FEP. D'autre part, dans [Rajagopalan et Chaudhuri, 1999] la scène et le flou sont modélisés en utilisant un modèle de champ de Markov. Ces deux inconnues sont ensuite estimées simultanément par MAP à l'aide d'une méthode de minimisation de type recuit simulé permettant d'éviter les minima locaux. Cette méthode est également utilisée dans [Rajan et Chaudhuri, 2003], mais dans cette référence, la carte de profondeur et la scène sont restaurées à une meilleure résolution que les images acquises. Enfin les références [Rajagopalan et Chaudhuri, 1998; Rajagopalan et Chaudhuri, 1997a; Rajagopalan *et al.*, 2000] sont plusieurs versions d'une même méthode d'estimation du flou de défocalisation sur plusieurs images par une technique de MV, basée sur une modélisation de l'image par un processus autorégressif à moyenne mobile (ARMA) et en utilisant la décomposition de l'image en Fourier et un modèle de FEP gaussiennes.

1.1.2.4 Modélisation du flou de défocalisation par un phénomène de diffusion

D'autres méthodes plus récentes modélisent la défocalisation par un phénomène de diffusion [Favaro *et al.*, 2004; Favaro *et al.*, 2008]. Dans [Favaro *et al.*, 2008] deux images ayant un flou variable sont utilisées, le principe étant d'estimer le flou relatif entre les deux images

en estimant la « quantité de diffusion » nécessaire pour passer de l'image la plus nette à l'image floue grâce à l'équation de la chaleur :

$$\begin{cases} \frac{\partial y}{\partial t}(m, n, t) = c \Delta y(m, n, t) \\ y(m, n, 0) = r(m, n) \end{cases} \quad (1.15)$$

r est l'image à l'état initial et la solution de cette équation $y(m, n, t)$ à un instant t correspond à l'une des images acquises. c est le coefficient de diffusion qui s'exprime en fonction de σ . Le lien entre le flou de défocalisation et les paramètres de l'équation de diffusion est donné par :

$$\sigma^2 = 2tc. \quad (1.16)$$

[Favaro *et al.*, 2004] étudient également le cas d'une image déformée à la fois par un flou de bougé et par un flou de défocalisation. Ils utilisent alors une équation de diffusion anisotrope en introduisant un tenseur de diffusion dans les équations (1.15). L'estimation des paramètres de flou de bougé, de la profondeur et de la scène est réalisée à l'aide d'une minimisation, par une descente de gradient, d'un critère suivant ces trois inconnues avec une régularisation de type Tikhonov.

1.1.2.5 Projection des données dans un sous-espace orthogonal

Dans [Soatto et Favaro, 2000] la convolution par la FEP est représentée par l'application d'un opérateur linéaire H_p sur la scène. La FEP est sélectionnée dans une famille de FEP potentielles par minimisation de la norme de la projection des données par le projecteur H_p^\perp :

$$\hat{p} = \underset{p}{\operatorname{argmin}} \|H_p^\perp \mathbf{Y}\|^2. \quad (1.17)$$

où \mathbf{Y} est la concatenation des données des différentes images et :

$$H_p^\perp = I - H_p H_p^+. \quad (1.18)$$

avec $^+$ correspondant à l'opérateur pseudo-inverse. H_p^\perp est un projecteur orthogonal qui projette les données sur le sous-espace orthogonal à $\operatorname{Im} H_p$. Dans cet article, la forme de la FEP n'est pas forcément gaussienne, mais simplement paramétrée en fonction de la profondeur. Les différentes matrices de projections H_p^\perp sont construites par une transformation des matrices résultantes de la décomposition en valeur singulière (SVD) de l'opérateur H_p . La méthode de [Soatto et Favaro, 2000] est reprise dans [Favaro et Soatto, 2005] mais une procédure de minimisation de (1.17) est proposée qui ne nécessite plus de connaître une expression paramétrée de la FEP mais qui repose sur un apprentissage. Le principe est de calculer la SVD d'une série d'images acquises à une profondeur p . Les vecteurs propres correspondant à des valeurs propres faibles constituent la base du sous-espace orthogonal à l'image de H_p . La minimisation de l'équation (1.17) revient alors à minimiser la norme de la projection des données sur les vecteurs de cette base.

1.1.2.6 Bilan

Les différentes méthodes d'estimation de profondeur multi-images reposent sur l'acquisition d'images d'une même scène suivant des configurations optiques différentes. Selon les modifications choisies, des étapes de recalage entre les deux images sont alors nécessaires, par exemple pour corriger les effets du grandissement ou d'une variation d'illumination. Plus

important, ces méthodes imposent que la scène soit statique entre les différentes acquisitions, ce qui limite le domaine d'application de ces méthodes. Une solution à ce problème est de séparer le flux incident. [Hiura et Matsuyama, 1998] utilisent par exemple un prisme dichroïque modifié et trois détecteurs CCD décalés et obtiennent ainsi trois images de mise au point différentes. Plus récemment, [Green *et al.*, 2007] divisent le faisceau de lumière incident à l'aide d'un ensemble de miroirs, ce qui permet d'obtenir plusieurs images acquises avec des ouvertures différentes. Par ailleurs, [Nagahara *et al.*, 2010] utilise un écran à cristaux liquides sur le trajet du faisceau, ce qui permet de modifier l'ouverture très rapidement entre deux acquisitions. Ces solutions entraînent néanmoins une augmentation de l'encombrement de l'appareil et de la difficulté de conception et de réalisation. Ces contraintes expérimentales ne sont pas rencontrées lors de l'estimation de FEP avec une seule image.

1.1.3 Estimation mono-image

Le cas mono-image consiste à estimer le flou à partir d'une seule image. Ce cas, que nous appelons SA-DFD pour *Single Acquisition Depth from Defocus*, est plus simple du point de vue expérimental, car une seule acquisition est nécessaire, mais en contrepartie la difficulté est reportée sur les traitements. En effet, contrairement au cas multi-images, l'estimation de FEP revient à vouloir résoudre une équation avec un jeu de deux inconnues. Dans les paragraphes suivants sont présentées les différentes méthodes décrites dans la littérature pour résoudre ce problème, en commençant par les méthodes d'estimation jointe de la scène et de la FEP, puis en présentant les méthodes qui utilisent un modèle de scène simplifié sur les contours de l'image, le phénomène de diffusion ou la sélection d'une FEP parmi une famille de FEP potentielles.

1.1.3.1 Utilisation des techniques de déconvolution aveugle

Le terme de déconvolution aveugle désigne l'estimation jointe de la scène et du flou à partir d'une seule acquisition. Dans [Kundur et Hatzinakos, 1996] sont présentées plusieurs de ces techniques et d'après les auteurs, les méthodes particulièrement adaptées à la photographie sont celles qui modélisent la scène par un processus autorégressif (AR). Par exemple, dans [Tekalp et Biemond, 1990], les paramètres du flou et de la scène, décrits par le modèle ARMA, sont estimés par une méthode de maximum de vraisemblance en utilisant simplement une hypothèse de flou symétrique. Dans [Reeves et Mersereau, 1992], le même formalisme est utilisé mais les auteurs emploient la validation croisée (VC) au lieu du maximum de vraisemblance. Plusieurs méthodes de déconvolution aveugle proposées récemment dans le cas de l'estimation du flou de bougé reposent sur l'utilisation d'a priori parcimonieux [Fergus *et al.*, 2006; Shan *et al.*, 2008; Cai *et al.*, 2009]. Plus précisément, dans [Shan *et al.*, 2008] des a priori parcimonieux sont utilisés pour modéliser l'image et le flou tandis que dans [Cai *et al.*, 2009] une représentation en *framelet* et *curvelet* est employée. Ces méthodes donnent des résultats très intéressants d'estimation du flou de bougé mais restent coûteuses en temps de calcul. Un processus plus rapide est proposé dans [Cho et Lee, 2009; Xu et Jia, 2010], en alternant entre l'estimation de la scène et l'estimation de la FEP.

Malgré les améliorations récentes des techniques de déconvolution aveugle, elles impliquent toujours des algorithmes complexes de minimisation à deux jeux d'inconnues. De plus, pour des critères dérivés d'un maximum a posteriori (MAP), l'estimation conjointe de la FEP et de la scène tend vers la solution triviale mais erronée où la scène est égale à l'image et le flou est un pic de Dirac [Fergus *et al.*, 2006; Levin *et al.*, 2009; Levin *et al.*, 2011]. Une solution proposée dans [Shan *et al.*, 2008] consiste à réajuster dans l'expression

du MAP, à chaque étape de l'optimisation, la pondération portant sur la probabilité de la scène, et celle portant sur le terme de vraisemblance. Cependant, cette solution augmente la complexité de l'algorithme. Une autre possibilité présentée dans [Kundur et Hatzinakos, 1996; Fergus *et al.*, 2006; Levin *et al.*, 2009] pour réduire la complexité des algorithmes et éviter la solution triviale consiste à estimer la FEP séparément de la scène.

1.1.3.2 Modélisation des contours de la scène par une fonction de Heaviside

Une première approche pour simplifier le problème de déconvolution aveugle consiste à modéliser les contours des objets de la scène par des fonctions de Heaviside orientées dans la direction des gradients comme l'illustre la figure 1.1(a). Avec l'utilisation d'un modèle de FEP gaussienne il est alors possible d'exprimer analytiquement le profil théorique de l'image floue. Grâce à ce modèle, une estimation du flou peut être obtenue en mesurant la distance entre les deux extrema de signes opposés de la dérivée seconde de l'image dans la direction du gradient [Elder et Zucker, 1998]. Cette technique, illustrée à la figure 1.1(b), dépend de la bonne estimation des extrema de la dérivée seconde, c'est pourquoi dans [Bae et Durand, 2007] cette distance est estimée par un critère des moindres carrés.

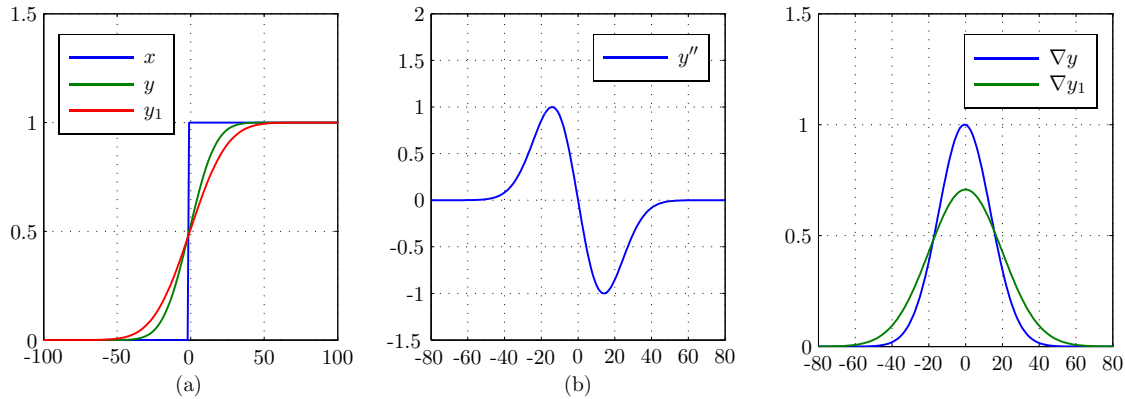


Fig. 1.1: (a) Modélisation de la scène par une fonction de Heaviside. (b) Principe de la méthode de [Elder et Zucker, 1998; Bae et Durand, 2007]. (c) Principe de la méthode de [Zhuo et Sim, 2009; Zhuo et Sim, 2011]

Par ailleurs, [Tai et Brown, 2009; Hu et De Haan, 2007; Tai *et al.*, 2009; Torreao et Fernandes, 2005; Zhuo et Sim, 2011; Zhuo et Sim, 2009; Cao *et al.*, 2011] génèrent une ou deux images supplémentaires par convolution de l'acquisition par une gaussienne de largeur fixe. Le flou relatif est alors estimé en utilisant le rapport des gradients des deux images au niveau du maximum comme l'illustre la figure 1.1(c). Par exemple dans [Zhuo et Sim, 2009; Zhuo et Sim, 2011], la valeur du flou s'obtient en calculant localement l'expression :

$$R = \frac{|\nabla y(0)|}{|\nabla y_1(0)|}, \quad (1.19)$$

où y est l'image acquise et y_1 l'image convoluée par une gaussienne de largeur σ_0 . La position 0 désigne la position théorique des maxima des gradients. La valeur de R permet d'estimer la valeur du flou par la relation :

$$\sigma = \frac{1}{\sqrt{R^2 - 1}} \sigma_0. \quad (1.20)$$

Ces méthodes sont fortement sensibles au bruit ainsi qu'à la position du maximum des gradients. C'est pourquoi [Bae et Durand, 2007; Zhang et Cham, 2009; Zhuo et Sim, 2011] proposent de lisser les résultats en utilisant un filtre bilatéral. En outre, ces méthodes donnent des cartes de profondeurs peu denses, car les résultats sont localisés uniquement sur les gradients. Cependant les résultats après régularisation sont de très bonne qualité, mais le traitement est alors très coûteux.

Une méthode basée sur les contours, plus générale que les méthodes précédentes, est celle de [Joshi *et al.*, 2008]. Joshi modélise également la scène au niveau des contours par une fonction de Heaviside orientée dans la direction perpendiculaire à la direction des gradients. Il estime localement les paramètres de cette marche puis s'en sert pour estimer la FEP par une technique de MAP. Ceci lui donne une coupe de la FEP dans la direction perpendiculaire au contour. Si l'image contient suffisamment de contours dans des directions variables, il en déduit la FEP. Cette méthode permet d'estimer localement uniquement une coupe de la FEP, il faut un ensemble suffisant de directions des contours pour estimer la FEP 2D. Cette méthode n'est donc pas pratique pour une estimation locale du flou de défocalisation, sauf si la FEP est paramétrique.

1.1.3.3 Modélisation du flou de défocalisation par un phénomène de diffusion

Comme dans le cas multi-image, le flou de défocalisation est parfois modélisé en utilisant le phénomène de diffusion [Namboodiri et Chaudhuri, 2008]. Dans cette référence, le flou est estimé en inversant le temps dans l'équation de la chaleur. Afin d'empêcher le résultat de cette inversion de diverger, les auteurs ajoutent un critère d'arrêt basé sur les gradients de l'image. L'inversion de l'équation de la chaleur s'arrête si la différence entre le gradient local et le gradient moyen local de l'image est trop importante.

Les méthodes citées précédemment reposent sur une hypothèse de FEP gaussienne ou paramétrique. Ces méthodes ne permettent donc pas de traiter le cas d'une optique non classique comme une optique à pupille codée [Veeraraghavan *et al.*, 2007; Levin *et al.*, 2007b; Zhou et Nayar, 2009a; Martinello *et al.*, 2010; Martinello et Favaro, 2011b].

1.1.3.4 Approches marginales

Une alternative aux méthodes de déconvolution aveugle consiste à baser l'estimation sur la densité de probabilité marginale de la FEP. La difficulté est alors de définir un modèle permettant cette marginalisation. Dans [Fergus *et al.*, 2006] les auteurs modélisent la distribution des gradients de la scène par un mélange de gaussienne à moyenne nulle. La probabilité jointe a posteriori des gradients de la scène \mathbf{x} et du flou k connaissant l'image \mathbf{y} , $p(\nabla\mathbf{x}, k | \nabla\mathbf{y})$, est estimée par le produit des distributions $q(k)q(\nabla\mathbf{y})$ suivant une approche variationnelle dans laquelle les distributions de k et des gradients de la scène sont supposées séparables. La FEP correspond à la moyenne de la distribution marginale $q(k)$ et l'image est ensuite déconvoluée à partir de cette FEP. Par ailleurs, les auteurs de [Levin *et al.*, 2009] proposent une méthode d'estimation de la FEP désignée par « MAP_k » où la scène est marginalisée dans le terme de vraisemblance $p(\mathbf{y} | k, \mathbf{x})$. Cette marginalisation est possible analytiquement uniquement pour un a priori simple sur la scène, comme l'a priori gaussien, c'est pourquoi Levin propose une méthode d'approximation de la densité de probabilité conditionnelle $p(\mathbf{x} | \mathbf{y}, k)$ par une approche variationnelle pour des a priori parcimonieux [Levin *et al.*, 2011]. L'estimation de la FEP est alors effectuée par une méthode de maximisation de l'espérance (EM). Notons que l'approche marginale est également proposée dans un contexte d'imagerie de la rétine [Blanco et Mugnier, 2011] ainsi que dans un

contexte multi-images [Blanc, 2002]. Dans ces références les auteurs illustrent notamment le gain de l'approche marginale vis-à-vis de l'estimation jointe des paramètres de la scène et de la FEP.

1.1.3.5 Sélection d'une FEP dans une famille de FEP potentielles

Une solution au problème d'estimation de FEP de forme quelconque, qui réduit fortement la complexité du traitement, consiste à sélectionner une FEP parmi une famille de FEP potentielles. Ces familles de FEP sont construites par calibrage de l'imageur ou par simulation en utilisant soit un logiciel de conception optique, si les plans de l'imageur sont connus soit des formules d'optique physique, telles que celles présentées dans l'annexe B. Plusieurs méthodes ont été développées pour traiter ce cas, notamment dans les articles utilisant une pupille codée pour favoriser l'estimation de profondeur.

Minimisation d'une erreur de reconstruction

Dans les références [Veeraraghavan *et al.*, 2007; Levin *et al.*, 2007b; Zhou et Nayar, 2009a] la profondeur est estimée localement avec un critère d'erreur de reconstruction. Le principe est de déconvoluer l'image avec toutes les FEP potentielles, de reconvoluer ces scènes estimées avec la FEP qui les a produites et de mesurer l'écart quadratique entre ce résultat et l'acquisition. Ceci revient à considérer le critère de minimisation suivant :

$$\hat{k} = \underset{k}{\operatorname{argmin}} \|y - h_k * x_k^{est}\|^2, \quad (1.21)$$

où x_k^{est} est le résultat de la déconvolution de l'image y par la FEP h_k . L'idée sous-jacente est que des artefacts de déconvolution tels que le ringing vont apparaître, dans les régions déconvoluées avec une mauvaise FEP, ce qui fait localement augmenter l'erreur de reconstruction. L'un des points délicats de ce type de critère est qu'il tend vers la solution triviale qui correspond à un flou de type Dirac et la scène égale à l'image acquise. Afin de corriger ce biais dans [Veeraraghavan *et al.*, 2007] le critère de reconstruction est régularisé en utilisant le kurtosis :

$$\hat{k} = \underset{k}{\operatorname{argmin}} \|y - h_k * x_k^{est}\|^2 - \beta \left(\text{kurtosis}(\nabla_v(x_k^{est})) + \text{kurtosis}(\nabla_h(x_k^{est})) \right), \quad (1.22)$$

où $\nabla_h(x)$ et $\nabla_v(x)$ correspondent respectivement aux gradients horizontaux et verticaux de x . Cependant ce critère nécessite un réglage du paramètre de régularisation β . Dans [Levin *et al.*, 2007b] les auteurs pondèrent le critère de reconstruction de la manière suivante :

$$\hat{k} = \underset{k}{\operatorname{argmin}} \lambda_k \sum_W \|y - h_k * x_k^{est}\|^2, \quad (1.23)$$

où W définit une fenêtre de l'image. Le réglage des paramètres λ_k est réalisé par un apprentissage sur des images à profondeur connue par une méthode non précisée.

Un autre défaut de ces approches est que le critère de reconstruction dépend de nombreux facteurs comme la méthode de déconvolution choisie, des paramètres utilisés pour effectuer cette déconvolution et les a priori choisis pour la scène, ce qui le rend difficilement reproductible sans toutes ces informations. Enfin, selon la méthode de déconvolution, les temps de calculs peuvent être très importants. Par exemple, Levin qui utilise un a priori non convexe obtient une déconvolution de l'image de taille 1600×1100 en près d'une heure, sur un CPU de 2.4Ghz.

Projection des données dans un sous-espace orthogonal obtenu par apprentissage

Dans [Martinello et Favaro, 2011a] et [Martinello et Favaro, 2011b], les auteurs utilisent la méthode proposée dans [Favaro et Soatto, 2005], décrite au paragraphe 1.1.2.5, dans le cas d'une seule image obtenue avec un imageur à pupille codée. À nouveau, une base de projection orthogonale associée à chaque profondeur est obtenue par apprentissage sur des images acquises à des profondeurs connues et la profondeur est estimée localement en minimisant la norme de la projection des données sur ces sous-espaces orthogonaux. Martinello et Favaro traitent également le cas d'un flou de bougé combiné à un flou de défocalisation [Martinello et Favaro, 2011a]. La famille de FEP potentielles est construite en combinant le flou de bougé (8 directions et 8 tailles potentielles) et le flou de défocalisation (10 profondeurs potentielles). La méthode proposée par Martinello donne des cartes de profondeur très propres, mais repose sur un apprentissage, ce qui limite la reproductibilité de la méthode ainsi que son potentiel pour la co-conception. Cependant, les essais de reproduction de cette méthode, présentés à la section 1.4.3 de ce chapitre ont conduit à des résultats très sensibles au bruit.

Approches bayésiennes non supervisées

Certaines approches de la littérature consistent à sélectionner la FEP parmi une famille de FEP potentielles en utilisant un formalisme bayésien et une approche par MV ou MAP. L'enjeu est à nouveau d'éliminer la dépendance vis-à-vis de la scène, qui est un paramètre inconnu. Un premier exemple est la référence [Klarquist *et al.*, 1995] dans laquelle la FEP est sélectionnée en maximisant l'expression obtenue en réinjectant l'estimateur du MV dans la vraisemblance. Dans le cas de l'estimation du flou de bougé, Levin propose d'utiliser la statistique des gradients de l'image [Levin, 2007]. Elle définit un ensemble de filtres 1D et modélise pour chaque filtre l'histogramme des gradients d'une image floue générée à partir de ces filtres. Puis elle compare ces histogrammes avec l'histogramme mesuré à partir des gradients de l'image acquise. Idéalement, il existe alors un filtre pour lequel les deux histogrammes se superposent, ce qui permet d'estimer le flou de défocalisation. Cette méthode permet de considérer plusieurs objets mobiles dans une image mais elle n'est adaptée qu'au cas du flou de bougé.

Dans ce même cadre, [Chakrabarti *et al.*, 2010] propose un critère de sélection de FEP à une dimension dans une famille de FEP potentielle en utilisant un critère de type maximum de vraisemblance après marginalisation. Ce critère est calculé après un filtrage des données pouvant s'interpréter comme une décomposition locale de Fourier à une dimension des données. Les gradients de la scène sont modélisés par un mélange de gaussienne et la variance du bruit est un paramètre fixé. Chakrabarti estime tout d'abord la direction et la taille du flou le plus probable dans toute l'image, en supposant un seul objet mobile, puis il construit une carte à deux labels objet mobile/fixe en maximisant la vraisemblance sur deux indices de FEP. Une régularisation spatiale tenant compte des différences de couleurs est ensuite appliquée par une technique de type *graphcut*.

Enfin, [Martinello *et al.*, 2010] est un exemple d'estimation de profondeur par MV marginale dans le cas d'un imageur à pupille codée. Les données sont tout d'abord préfiltrées par un filtre passe-haut et la covariance de la scène filtrée est supposée diagonale. Martinello obtient ainsi une expression de la vraisemblance marginale de forme gaussienne, dont la covariance dépend de la structure de la FEP. L'hypothèse utilisée ensuite par Martinello est que la covariance des données dérive d'un champ de Markov. La vraisemblance marginale est alors réécrite en utilisant une pseudovraisemblance faisant intervenir uniquement un voisinage dépendant de la structure de la pupille codée et de l'hypothèse de profondeur.

Finalement, le critère de vraisemblance est régularisé par une pénalisation sur la valeur seuillée de l'écart entre les pixels voisins et par un *graphcut*. L'inconvénient majeur de cette méthode est que le voisinage et donc la complexité de l'algorithme dépend du nombre de trous de la pupille.

1.1.4 Bilan

Parmi les différentes méthodes que nous avons présentées, les méthodes d'estimation du flou de défocalisation reposant sur l'utilisation de plusieurs images ont de fortes contraintes expérimentales, car elles nécessitent une scène statique durant les acquisitions, ou impliquent un encombrement important de l'imageur. Le cas mono-image ne présente pas cette difficulté expérimentale mais il est plus complexe du point de vue traitement, car il s'agit alors d'estimer deux jeux d'inconnues sur une seule image. D'un côté, les techniques classiques de déconvolution aveugle conduisent à des traitements lourds. D'un autre côté, les méthodes reposant sur une estimation du flou sur les contours ou sur la modélisation de l'image par un phénomène de diffusion, contraignent la forme de la FEP et ne permettent pas de traiter les pupilles codées. De plus, les méthodes basées sur les contours ne favorisent pas un résultat dense, sauf lorsqu'elles sont suivies d'une régularisation coûteuse qui étendent les résultats d'estimation de profondeur sur l'ensemble de l'image. Les méthodes permettant de traiter les formes de FEP les plus générales sont celles qui sélectionnent une FEP parmi une famille de FEP potentielles. Plusieurs méthodes proposées dans la littérature reposent sur une étape d'apprentissage sur une base d'images. Or, comme annoncé dans l'introduction, l'apprentissage est un obstacle à la co-conception pour des raisons de reproductibilité et de risque de surapprentissage. Enfin, les méthodes traitant de la sélection non supervisées de la FEP dans une famille de FEP potentielles sont soit dédiées au cas du flou de bougé, soit ont une complexité qui dépend de la forme de la FEP. En conclusion, il n'existe pas dans la littérature de méthode non supervisée d'estimation de FEP de forme quelconque. Ceci motive le développement de l'algorithme présenté dans la suite de ce chapitre.

1.2 Algorithme d'estimation de FEP dans le cas mono-image

Dans la suite de ce chapitre nous présentons l'algorithme d'estimation locale de FEP développé au cours de cette thèse, nommé GLDFD pour *Generalized Likelihood Depth from Defocus*. Afin de pouvoir être utilisé avec une large gamme d'optiques, notre algorithme sélectionne localement une FEP dans une famille de FEP potentielles obtenues par calibrage ou par simulation. La sélection de la FEP est réalisée à l'aide d'un critère dérivé d'une technique de maximisation de la vraisemblance généralisée obtenue en intégrant la scène « hors du problème ». Pour cela, nous utilisons un a priori gaussien isotrope pour modéliser les gradients de la scène. Cet a priori a la particularité d'être impropre, autrement dit la matrice de covariance de la scène n'est pas définie. La définition d'une vraisemblance marginale des données de l'image dans ce cas est délicate, mais nous montrons à l'aide des propriétés présentées dans l'annexe A, extraites du rapport technique de F. Champagnat, qu'une transformation des données permet toutefois de définir cette vraisemblance en fonction de la matrice de précision notée Q , représentant l'inverse de la matrice de covariance [Champagnat, 2012]. Le critère que nous proposons, nommé GL, dépend finalement de deux paramètres qui sont le rapport signal à bruit et la FEP. Ces deux paramètres sont estimés de manière non supervisée sur chaque fenêtre de l'image. Une implémentation efficace de ce critère est également proposée dans ce chapitre, qui permet une estimation de la FEP sur des fenêtres de tailles réduites de l'image sans faire d'hypothèse sur les bords de l'image.

Dans cette section, l'algorithme GLDFD est décrit dans le cas mono-image, mais une généralisation au cas multi-images est possible, comme nous le montrons dans le chapitre 2.

1.2.1 Modélisation des données

Comme discuté dans la section 1.1.1, l'estimation de la FEP doit être locale car, pour l'estimation du flou de bougé comme du flou de défocalisation, la FEP varie spatialement dans l'image. C'est pourquoi la FEP est supposée constante seulement sur des régions réduites de l'image. La relation locale entre l'image et la scène, sous forme matricielle, est donnée par la relation suivante :

$$\mathbf{y} = H_k \mathbf{x} + \mathbf{b}, \quad (1.24)$$

où \mathbf{y} contient les N pixels appartenant à une fenêtre de l'image dans la représentation lexicographique, \mathbf{x} les M pixels de la scène et \mathbf{b} est un vecteur aléatoire modélisant le bruit d'acquisition. H_k est la matrice de convolution $N \times M$ relative à la FEP k de la famille $\{H_1, \dots, H_K\}$. Comme les fenêtres considérées sont de taille réduite, les effets de bords ne sont pas négligeables, c'est pourquoi la matrice de convolution utilisée ici est de type « valide » autrement dit elle n'implique aucune approximation sur les bords. Ceci revient à considérer que le support de \mathbf{x} est plus grand que celui de \mathbf{y} [Idier, 2001, Section 4.3.2]. Pour simplifier les expressions dans la suite de cette section, la matrice de convolution sera notée H .

Le bruit s'ajoutant à la scène convoluée par la FEP est modélisé par un vecteur aléatoire blanc gaussien de moyenne nulle et de matrice de précision, inverse de la matrice de covariance, notée Q_b . La densité de probabilité de l'image connaissant la scène s'écrit alors :

$$p(\mathbf{y}|\mathbf{x}, H, Q_b) = (2\pi)^{-N/2} |Q_b|^{1/2} \exp - \frac{(\mathbf{y} - H\mathbf{x})^T Q_b (\mathbf{y} - H\mathbf{x})}{2}, \quad (1.25)$$

où $|\cdot|$ désigne le déterminant.

Nous proposons d'estimer localement la FEP parmi une famille de FEP potentielles en maximisant la vraisemblance marginale des données :

$$p(\mathbf{y}|H, Q_b, Q_x) = \int p(\mathbf{y}|\mathbf{x}, H, Q_b) p(\mathbf{x}|Q_x) d\mathbf{x}. \quad (1.26)$$

Pour obtenir une expression analytique de la vraisemblance marginale, il est donc nécessaire de modéliser la densité de probabilité de la scène.

1.2.2 Modélisation de la scène

Plusieurs méthodes de déconvolution aveugle récentes utilisent un a priori parcimonieux dit « naturel » qui favorise une image constante par morceaux [Levin *et al.*, 2007a; Levin *et al.*, 2007b; Levin *et al.*, 2009]. Cependant, ces a priori rendent impossible le calcul analytique de la vraisemblance marginale, que nous souhaitons utiliser pour estimer la FEP. En revanche, l'a priori gaussien permet ce calcul et peut conduire à des résultats d'estimation de FEP acceptables [Levin *et al.*, 2009]. C'est pourquoi nous proposons d'utiliser une densité de probabilité de la scène de la forme :

$$p(\mathbf{x}|Q_x) \propto \exp - \frac{\mathbf{x}^T Q_x \mathbf{x}}{2}, \quad (1.27)$$

où Q_x est une matrice de précision associée à la scène.

Dans les paragraphes suivants sont données les expressions de la vraisemblance marginale

dans deux cas. Tout d'abord lorsque la matrice de précision est régulière, cas classique désigné par le terme d'a priori *propre*, puis dans le cas où la matrice de précision est singulière, cas désigné par le terme d'a priori *impropre*.

1.2.3 Vraisemblance marginale dans le cas d'un a priori propre

Lorsque la matrice de précision Q_x est régulière, l'a priori (1.27) s'écrit :

$$p(\mathbf{x}|Q_x) = (2\pi)^{-M/2} |Q_x|^{1/2} \exp - \frac{\mathbf{x}^t Q_x \mathbf{x}}{2}. \quad (1.28)$$

Dans ce cas, Q_x est simplement l'inverse de la matrice de covariance et le terme à l'intérieur de l'intégrale dans (1.26) correspondant à la probabilité jointe des données et de la scène s'écrit :

$$(2\pi)^{-(N+M)/2} |Q_b|^{1/2} |Q_x|^{1/2} \exp - \frac{1}{2} (\mathbf{x}^t Q_x \mathbf{x} + (\mathbf{y} - H\mathbf{x})^t Q_b (\mathbf{y} - H\mathbf{x})) \quad (1.29)$$

En effectuant le changement de variable $[\mathbf{x}] \rightarrow [\mathbf{x}_2]$ avec $\mathbf{x}_2 = \mathbf{x} - \hat{\mathbf{x}}$, et $\hat{\mathbf{x}}$ donné par l'estimation de \mathbf{x} par MAP :

$$\hat{\mathbf{x}} = (H^t Q_b H + Q_x)^{-1} H^t Q_b \mathbf{y}, \quad (1.30)$$

l'argument de l'exponentielle à l'intérieur de l'intégrale dans (1.26) devient :

$$\mathbf{x}^t Q_x \mathbf{x} + (\mathbf{y} - H\mathbf{x})^t Q_b (\mathbf{y} - H\mathbf{x}) = \mathbf{x}_2^t Q_{x_2} \mathbf{x}_2 + \mathbf{y}^t Q_y \mathbf{y} \quad (1.31)$$

où

$$Q_{x_2} = H^t Q_b H + Q_x \quad (1.32)$$

$$Q_y = Q_b - Q_b H (H^t Q_b H + Q_x)^{-1} H^t Q_b. \quad (1.33)$$

En utilisant l'équation (1.29), il est alors possible de calculer analytiquement la vraisemblance marginale.

L'équation (1.33) donne l'expression de la matrice de précision associée à \mathbf{y} , Q_y . Cette matrice dépend de la FEP et des deux matrices de précision Q_x et Q_b . En utilisant la notation générale $\boldsymbol{\theta} \triangleq \{H, Q_b, Q_x\}$, pour désigner les différents paramètres dont dépend Q_y , le calcul de l'intégrale (1.26) permet d'obtenir une expression analytique de la vraisemblance marginale notée classiquement $L(\mathbf{y}|\boldsymbol{\theta})$:

$$L(\mathbf{y}|\boldsymbol{\theta}) = p(\mathbf{y}|H, Q_b, Q_x) = \left| \frac{Q_y(\boldsymbol{\theta})}{2\pi} \right|^{\frac{1}{2}} \exp - \frac{1}{2} (\mathbf{y}^t Q_y(\boldsymbol{\theta}) \mathbf{y}). \quad (1.34)$$

1.2.4 Vraisemblance marginale dans le cas d'un a priori impropre

Nous choisissons de définir la matrice de précision en fonction des dérivées du premier ordre de la scène. Cet a priori est couramment utilisé dans la littérature dans les termes de régularisation pour des estimation de type MAP. Ainsi, nous définissons la matrice de précision de la scène par :

$$Q_x = \frac{D^t D}{\sigma_x^2}, \quad (1.35)$$

σ_x^2 caractérise la variance des gradients de la scène et D est la concaténation verticale des deux matrices de convolution L_1 et L_2 respectivement associées aux opérateurs de

dérivées horizontales $[-1 \ 1]$ et verticales $[-1 \ 1]^t$. Notons que D est alors de dimension $2(M - \sqrt{M}) \times M$. Le problème est qu'avec cet a priori, la matrice Q_x n'est pas régulière. En effet, le vecteur $\mathbf{1}$ est un vecteur du noyau de Q_x . La définition d'une vraisemblance telle que (1.34) n'est plus possible car le déterminant de Q_x est nul. Dans [Orieux *et al.*, 2010], cette difficulté est contournée en ajoutant un terme dans l'a priori selon :

$$p(\mathbf{x}|\sigma_x^2, \beta) \propto \exp\left(-\frac{\|D\mathbf{x}\|^2}{2\sigma_x^2} + \beta\|\mathbf{x}\|^2\right), \quad (1.36)$$

mais cette astuce ajoute le paramètre β au problème, paramètre qui doit alors être optimisé pour obtenir des résultats satisfaisants.

La manipulation d'a priori impropres est délicate et il existe peu de références présentant les propriétés de ces modèles, à l'exception de [Besag et Kooperberg, 1995]. Nous utilisons ici un résultat de F. Champagnat rappelé dans l'annexe A de ce manuscrit. En effet, d'après la Proposition 1 de l'annexe A, il est possible de définir la vraisemblance d'un paramètre θ associé à un vecteur aléatoire \mathbf{y} ayant une matrice de précision Q_θ singulière, grâce à une transformation linéaire de ce vecteur séparant la composante propre et la composante impropre. La composante impropre ne dépendant pas de θ , la vraisemblance de θ est définie en utilisant celle associée à la composante propre. D'après la Proposition 1, la vraisemblance associée au paramètre θ peut alors s'écrire :

$$L(\mathbf{y}|\theta) = \left| \frac{Q_y(\theta)}{2\pi} \right|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{y}^t Q_y(\theta) \mathbf{y}\right). \quad (1.37)$$

où $|Q_y(\theta)|_+$ est le produit des valeurs propres non nulles de $Q_y(\theta)$. La condition à respecter est que le noyau de Q_θ soit indépendant de θ .

Dans le cas de l'estimation de profondeur et avec l'utilisation de la matrice Q_x définie dans l'équation (1.35), la scène est un vecteur aléatoire gaussien impropre. D'après la section A.4 de l'annexe A, le vecteur \mathbf{y} est également un vecteur aléatoire gaussien dont la matrice de précision s'écrit :

$$Q_y(\theta) = Q_b - Q_b H (H^t Q_b H + Q_x)^{-1} H^t Q_b, \quad (1.38)$$

La proposition 1 peut s'appliquer à \mathbf{y} avec $\theta \triangleq \{H, Q_b, Q_x\}$, à condition que le noyau de $Q_y(\theta)$ soit indépendant de θ . D'après le lemme 4 de l'annexe A : $\text{Ker } Q_y(\theta) = H \text{Ker } (Q_x)$. D'où :

$$\text{Ker } Q_y(\theta) = H \text{Ker } (D^t D) = H \text{Ker } D. \quad (1.39)$$

Puisque D est une concaténation des opérateurs de dérivation horizontale et verticale du premier ordre, son noyau se réduit au cas de scènes constantes, soit $\text{Ker } D = \{\mathbf{1}\}$. Sachant que H est une matrice de convolution alors $H\mathbf{1} \propto \mathbf{1}$. Ainsi, le noyau de $Q_y(\theta)$ ne dépend pas de θ , la condition de la proposition 1 est donc respectée, ce qui permet d'utiliser l'équation (1.37).

1.2.5 Dérivation d'un critère d'estimation de FEP

L'expression de la vraisemblance marginale des données, obtenue à partir d'un a priori impropre, nous permet de définir un critère d'estimation de profondeur. En effet, en supposant que le bruit est blanc gaussien et centré, sa matrice de précision est $Q_b = I/\sigma_b^2$ avec I

la matrice identité. De plus, Q_x étant donnée par (1.35), la matrice de précision des données \mathbf{y} s'écrit :

$$Q_y(\boldsymbol{\theta}) = \frac{1}{\sigma_b^2} (I - H(H^t H + \alpha D^t D)^{-1} H^t), \quad (1.40)$$

où le paramètre de régularisation $\alpha = \sigma_b^2 / \sigma_x^2$ peut s'interpréter comme l'inverse d'un rapport signal à bruit. D'après le paragraphe précédent, la vraisemblance marginale peut s'écrire :

$$L(\mathbf{y}|\boldsymbol{\theta}) = (2\pi)^{-\left(\frac{N-1}{2}\right)} |Q_y(\boldsymbol{\theta})|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{y}^t Q_y(\boldsymbol{\theta}) \mathbf{y}\right). \quad (1.41)$$

$\boldsymbol{\theta}$ est alors défini par $\{H, \sigma_b^2, \alpha\}$. Il reste cependant toujours trois paramètres à estimer. Nous proposons donc de réduire le nombre de paramètres en maximisant la vraisemblance vis-à-vis de σ_b^2 afin de travailler avec une *vraisemblance généralisée* qui dépend seulement de H et de α [Idier, 2001, Section 3.8.2]. Ainsi, la matrice de précision est réécrite en séparant le paramètre σ_b^2 , soit $Q_y(\boldsymbol{\theta}) = \sigma_b^{-2} P_\psi$, où

$$P_\psi = I - H(H^t H + \alpha D^t D)^{-1} H^t, \quad (1.42)$$

et $\psi \triangleq \{H, \alpha\}$. La vraisemblance (1.37) s'écrit alors :

$$L(\mathbf{y}|\psi, \sigma_b^2) = (2\pi\sigma_b^2)^{-\left(\frac{N-1}{2}\right)} |P_\psi|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{y}^t \frac{P_\psi}{\sigma_b^2} \mathbf{y}\right). \quad (1.43)$$

La maximisation de la vraisemblance par rapport à σ_b^2 donne : $\hat{\sigma}_b^2 = \mathbf{y}^t P_\psi \mathbf{y} / (N - 1)$. Ce terme est alors réintroduit dans (1.43) ce qui implique :

$$L(\mathbf{y}|\psi, \hat{\sigma}_b^2) \propto |P_\psi|_+^{\frac{1}{2}} (\mathbf{y}^t P_\psi \mathbf{y})^{-\left(\frac{N-1}{2}\right)}.$$

Maximiser la vraisemblance revient à minimiser le critère GL, faisant référence au terme de *generalized likelihood* utilisé dans [Wahba, 1985], qui est défini par :

$$\text{GL}(\psi) = \mathbf{y}^t P_\psi \mathbf{y} |P_\psi|_+^{\frac{-1}{(N-1)}} \quad (1.44)$$

L'objectif est ici d'identifier localement une FEP H_k parmi une famille de FEP potentielles $\{H_1, \dots, H_K\}$. En notant $\hat{\alpha}_k = \arg \min_{\alpha} \text{GL}(H_k, \alpha)$ nous proposons d'utiliser le critère de sélection suivant :

$$\hat{k} = \arg \min_k \text{GL}(H_k, \hat{\alpha}_k). \quad (1.45)$$

1.2.6 Description de l'algorithme GLDFD

Le traitement consiste à décomposer l'image en un ensemble de fenêtres, avec ou sans recouvrement. Pour chaque fenêtre, la FEP k est déterminée par une double minimisation représentée à la figure 1.2. Pour une valeur de k fixée, le critère GL est optimisé par une minimisation continue à une dimension sur α , effectuée avec la fonction `fminbnd` de Matlab qui repose sur la méthode de minimisation appelée « golden section search ». Les bornes pour la recherche de α sont 10^{-6} et 1. Puis le critère $\text{GL}(H_k, \hat{\alpha}_k)$ est minimisé suivant un ensemble discret de valeurs de k , ce qui permet d'obtenir \hat{k} . Notons que l'estimation de k n'est pas effectuée sur les régions homogènes qui sont insensibles à la FEP. Comme illustré dans le schéma de la figure 1.2, ces régions sont rejetées préalablement par un test de la texture fondé un détecteur de Canny. L'algorithme ainsi obtenu est désigné par le terme GLDFD.

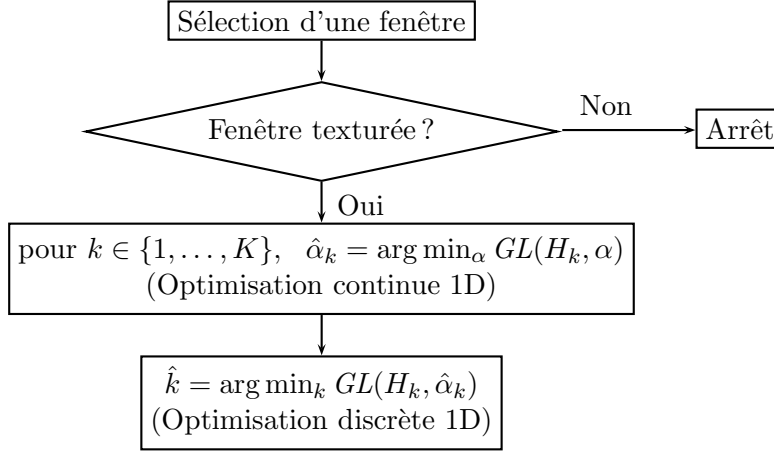


Fig. 1.2: Schéma bloc de l'algorithme d'estimation de profondeur GLDFD.

1.2.7 Exemple illustratif 1D

Dans cet exemple la famille de FEP potentielles est constituée de trois FEP rectangulaires 1D de longueur 3, 6 et 9 pixels. Dans un premier temps, la scène \mathbf{x} correspond à une fonction porte de largeur 12 pixels. Le signal \mathbf{y} présenté à la figure 1.3 (a) est simulé par convolution de cet objet avec la PSF de largeur 6 pixels et l'addition d'un bruit blanc gaussien d'écart-type 0.05. La figure 1.3 (b) montre les courbes du logarithme du critère GL en fonction de α pour ce signal et les trois FEP de la famille de FEP potentielles. Le minimum global du critère GL correspond bien à la FEP de largeur 6 pixels. Cependant, la courbe associée à cette FEP n'est pas minimale pour toutes les valeurs de α , d'où la nécessité de chercher pour chaque FEP la valeur de α qui minimise le critère GL puis de comparer les minima entre eux pour sélectionner une FEP. Cet exemple justifie donc l'approche décrite dans le schéma bloc 1.2. Ensuite, la scène \mathbf{x} considérée est homogène et le signal, présenté à la figure 1.3 (b), est obtenu par convolution de \mathbf{x} avec la FEP de longueur 6 pixels et l'ajout d'un bruit blanc gaussien d'écart-type 0.05. La figure 1.3 (c) présente les courbes du logarithme du critère GL en fonction de α pour les trois FEP. Dans ce cas, il n'y a aucun minimum sur les courbes de GL et toutes les FEP sont alors équivalentes. Ce résultat justifie le rejet des régions homogènes insensibles à la FEP dans une étape de prétraitement.

1.2.8 Implémentation

Le calcul direct du critère GL défini par l'équation (1.44) requiert un temps de calcul important à cause des dimensions de la matrice $H^t H + \alpha D^t D$. Typiquement, pour une fenêtre de données et une FEP de même taille 21×21 pixels, cette matrice est de taille 1681×1681 pixels et doit être inversée à chaque optimisation de α . Classiquement les matrices $H^t H$ et $D^t D$ sont diagonalisées dans la base de Fourier qui fait l'hypothèse d'une scène périodique. Or cette approximation n'est pas valide pour des fenêtres dont la taille est du même ordre de grandeur que la FEP, car elle introduit des effets de bords. La solution que nous proposons consiste à effectuer une décomposition de ces matrices en valeurs singulières généralisées (GSVD) [Neumaier, 1998]. Pour deux matrices H de dimension $N \times M$ et D

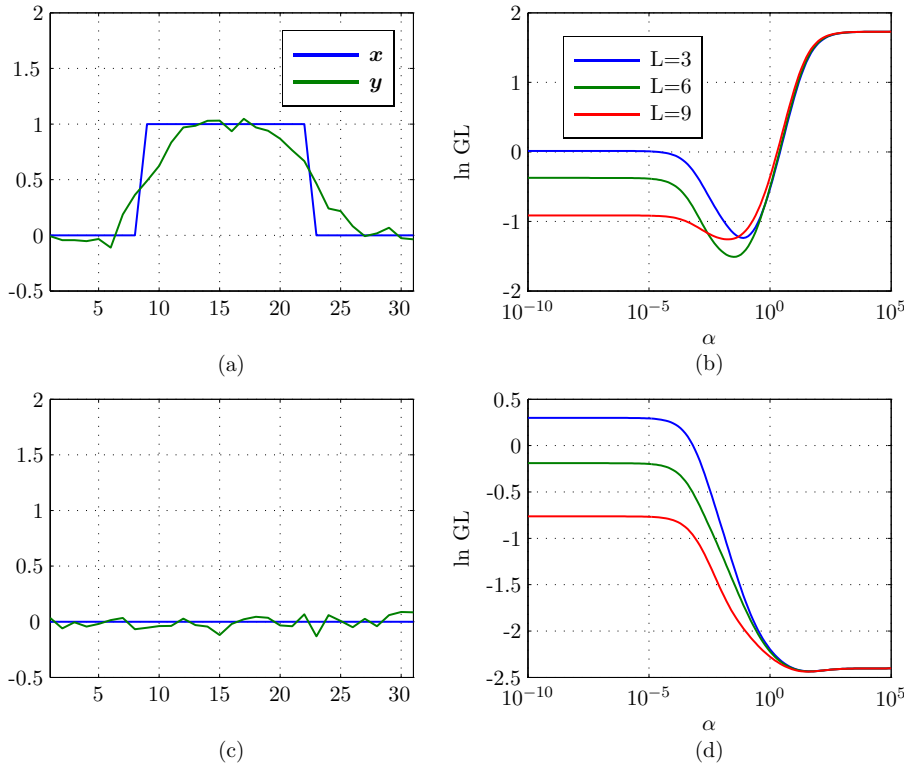


Fig. 1.3: Exemple de courbes du critère GL en fonction de α pour trois FEP dans le cas d'une scène constituée d'une marche ((a) et (b)) ou d'une scène uniforme ((c) et (d)).

de dimension $P \times M$ la GSVD s'écrit :

$$H = UCX^t \quad (1.46)$$

$$D = VSX^t \quad (1.47)$$

$$I = C^tC + S^tS, \quad (1.48)$$

où U (respectivement V) est une matrice unitaire de taille $N \times N$ (resp. $P \times P$), I est la matrice identité. Les matrices C et S sont respectivement de taille $N \times r$ et $P \times r$ où r est le rang de la concaténation verticale des matrices H et D . Comme illustré dans l'annexe D, les matrices C^tC et S^tS sont des matrices diagonales. Avec cette décomposition, en remplaçant (1.46) et (1.47) dans (1.42), la matrice P_ψ devient :

$$P_\psi = I - UC(C^tC + \alpha S^tS)^{-1}C^tU^t, \quad (1.49)$$

$$= U\Omega U^t, \quad (1.50)$$

avec $\Omega = I - C(C^tC + \alpha S^tS)^{-1}C^t$. Dans l'annexe D, nous montrons que le critère GL peut alors s'écrire :

$$\text{GL}(\psi) = \left(\prod_{i=1}^{N-1} \omega_i \right)^{-\frac{1}{N-1}} \left(\sum_{i=1}^N \omega_i z_i^2 \right), \quad (1.51)$$

où avec $z = U^t y$ et $\omega_i = \frac{\alpha s_i^2}{c_i^2 + \alpha s_i^2}$, où c^2 et s^2 correspondent respectivement aux N dernières valeurs diagonales de C^tC et S^tS et par convention $s_N=0$. L'avantage de la décomposition en GSVD est que les matrices U , V , C , S et X sont indépendantes de α . Ainsi pour un

imageur donné, il est possible de calculer ces matrices dans une étape de prétraitement et de les stocker, pour traiter ensuite n'importe quelle image issue de cet imageur. En pratique seuls les coefficients diagonaux de la matrice C et la matrice U sont utilisés car S vérifie $S^t S = I - C^t C$ et les matrices V et X n'interviennent pas dans le calcul du critère GL.

1.2.9 Bilan

Dans les paragraphes précédents, nous avons décrit un nouveau critère d'estimation locale de flou qui sélectionne une FEP parmi une famille de FEP potentielles. Cet algorithme repose sur une modélisation de la scène par un a priori gaussien isotrope impropre et nous avons montré comment exprimer une vraisemblance des données avec ce type d'a priori. Le critère GL que nous proposons dépend de deux paramètres qui sont le rapport signal à bruit et la FEP. Ces deux paramètres sont estimés de manière non supervisée sur chaque fenêtre de l'image. Dans la suite de ce chapitre, l'algorithme GLDFD est testé sur des données simulées afin d'étudier sa robustesse vis-à-vis des conditions d'utilisation, comme le niveau de bruit, le nombre de données ou la texture de la scène. Puis il est comparé à plusieurs algorithmes de la littérature dédiés au flou de bougé et au flou de défocalisation sur des images simulées et réelles.

1.3 Validation de l'algorithme GLDFD sur des données simulées

L'objectif de cette section est d'étudier les performances de l'algorithme présenté dans la section précédente sur des images simulées dans différentes conditions d'utilisation, en faisant varier par exemple le nombre de données, le niveau de bruit ou la texture de la scène. Les simulations sont effectuées dans le cas de l'estimation du flou de défocalisation pour un système optique idéal, autrement dit sans aberration autre que la défocalisation, dont les FEP sont simulées pour différentes profondeurs en utilisant les formules d'optique de Fourier décrites dans l'annexe B. Le tableau 1.1 présente les caractéristiques du système optique utilisées pour simuler la famille de FEP.

Focale	25 mm
Ouverture $F/\#$	2.5
p_x	5 μm
p_0	2 m
Taille de la FEP	21 \times 21

Tableau 1.1: Paramètres de l'optique et du détecteur utilisés pour simuler les FEP d'un imageur idéal.

Pour chaque profondeur, les FEP sont simulées pour trois longueurs d'onde du visible choisie arbitrairement à 470, 550 et 640 nm, car ces longueurs d'onde correspondent aux maxima classiques des spectres des filtres bleu, vert et rouge d'un détecteur couleur. Puis ces FEP sont additionnées afin de simuler une FEP polychromatique. Les profondeurs considérées dans ces simulations sont comprises entre 2 m et 3 m avec un pas entre chaque FEP de 20 mm. Tout d'abord l'algorithme GLDFD est testé sur des scènes vérifiant l'a priori gaussien, générées en utilisant la méthode décrite dans l'annexe C, puis sur des scènes naturelles extraites de photographies tirées d'Internet. Dans la plupart des simulations suivantes, une image est générée en concaténant horizontalement quinze fenêtres de taille 21 \times 21, chacune construite par la convolution d'une scène avec une même FEP puis l'ajout d'un bruit blanc

gaussien d'écart-type σ_b . Ces fenêtres, relatives à une FEP, sont alors concaténées verticalement pour différentes FEP de la famille, soit différentes valeurs de profondeur. Notons que les scènes ont un support supérieur à l'image, et nous conservons uniquement les pixels de l'image issus d'une convolution de type « valide » avec la FEP.

1.3.1 Scènes respectant l'a priori gaussien

Dans les simulations suivantes la scène de chaque scène est simulée en utilisant la méthode décrite dans l'annexe C. Le paramètre σ_x^2 est fixé arbitrairement à 50, et la valeur de σ_b est adaptée pour différentes valeur de $\alpha = \sigma_b^2 / \sigma_x^2$.

1.3.1.1 Estimation de la FEP

L'objectif de ce premier exemple est de tester les performance d'estimation de FEP de l'algorithme GLDFD sur un ensemble de profondeurs avec différents niveaux de rapport signal à bruit. La figure 1.4 présente les résultats d'estimation de profondeur en utilisant le critère GL pour α valant successivement 10^{-4} , 10^{-3} , 10^{-2} et 10^{-1} . Pour de faibles valeurs de

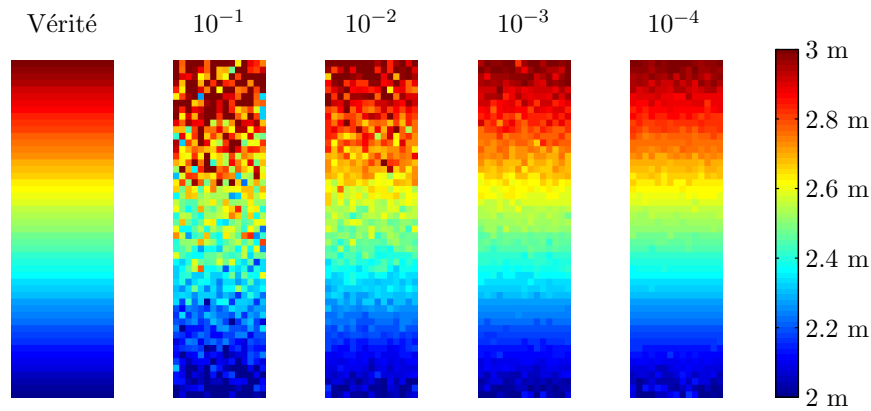


Fig. 1.4: Résultats d'estimation de profondeur avec le critère GL dans le cas de scènes vérifiant l'a priori gaussien et pour différentes valeurs de α pouvant s'interpréter comme l'inverse du rapport signal à bruit.

α , donc pour un rapport signal à bruit élevé, l'estimation de profondeur donne des résultats satisfaisants où l'on retrouve presque exactement la vérité. Lorsque le bruit augmente, la variance des résultats augmente mais les valeurs de profondeur restent proches de la vérité. Ce test confirme la capacité d'estimation de FEP de l'algorithme GLDFD lorsque la scène respecte l'a priori gaussien utilisé pour exprimer le critère GL.

1.3.1.2 Estimation du paramètre α

Le critère (1.45) prévoit l'optimisation jointe de deux paramètres, l'indice de la FEP et le paramètre de régularisation α . Dans cet exemple, l'objectif est de tester l'estimation de α par rapport à la valeur théorique fixée lors de la génération de chaque fenêtre. En reprenant les résultats du test précédent, pour chaque valeur de α , la moyenne des $\hat{\alpha}$ retenus par l'algorithme GLDFD est calculée sur l'ensemble des fenêtres. La figure 1.5 représente la moyenne et l'écart-type des résultats d'estimation de α sur l'ensemble des profondeurs comprises entre 2.2 et 3 m, en fonction de la valeur α théorique.

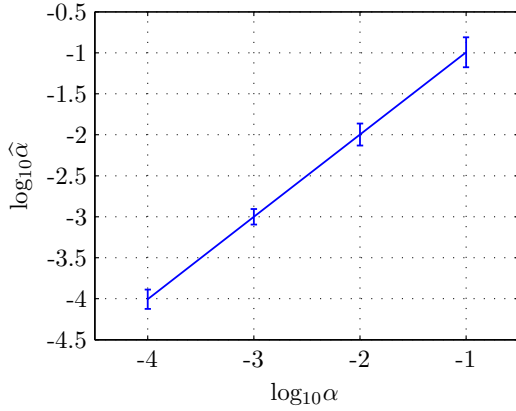


Fig. 1.5: Résultats de l'estimation de α par l'algorithme GLDFD.

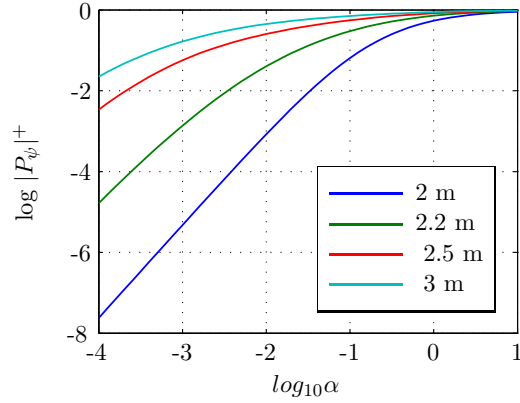


Fig. 1.6: Variation du dénominateur du critère GL en fonction de α .

Les moyennes des estimations de α correspondent aux valeurs théoriques. L'écart-type sur les valeurs estimées de $\log_{10}\alpha$ est compris entre 0.1 et 0.2. Ce test confirme la capacité d'estimation du paramètre α avec le critère GL. Notons qu'ici nous avons fait varier le paramètre α en modifiant la variance du bruit mais des résultats similaires sur l'estimation de α sont obtenus en fixant la variance du bruit et en faisant varier la variance des gradients de la scène.

1.3.1.3 Variations du dénominateur du critère GL pour différentes FEP

Le point délicat lors de l'estimation aveugle de la FEP est d'éviter la solution triviale qui consiste à avoir $H = I$ et $\mathbf{y} = \mathbf{x}$ [Fergus *et al.*, 2006; Levin *et al.*, 2009]. Dans le critère GL un biais vers les FEP les plus étroites, correspondant au plan de mise au point, est à craindre car le numérateur peut se mettre sous la forme $\mathbf{y}^t(\mathbf{y} - H\mathbf{x})$. Cependant, lors des simulations de la section 1.3.1.1, les FEP allant jusqu'à la position du plan de mise au point, à savoir 2 m, ont été utilisées, et aucun biais n'a été observé. En effet, dans le critère GL le numérateur est pondéré par l'inverse du déterminant de la matrice P_ψ . Le logarithme de ce terme est représenté à la figure 1.6 pour quelques profondeurs en fonction du logarithme de α . Les valeurs les plus faibles du dénominateur correspondent aux profondeurs proches du plan de mise au point à 2 m, *i.e.*, aux FEP les plus étroites. Ainsi, ces FEP sont plus fortement pénalisées que les FEP larges. C'est donc la pondération par l'inverse du déterminant de P_ψ qui permet d'éviter le biais vers les FEP les plus étroites.

1.3.1.4 Variation du nombre de données

L'objectif de cette simulation est d'étudier l'influence du nombre de données, autrement dit de la taille de la fenêtre, sur les résultats d'estimation de profondeur. Dans cette simulation, les tailles de fenêtres considérées sont les tailles : 3×3 , 7×7 , 15×15 , 21×21 et 35×35 pixels. Pour une profondeur donnée et pour chaque taille, 144 fenêtres sont générées à partir de scène respectant l'a priori gaussien telles que $\alpha = 0.001$ et l'algorithme GLDFD estime la profondeur sur ces fenêtres. Cette opération est répétée pour plusieurs profondeurs : 2.8 m, 2.6 m, 2.4 m, 2.2 m et 2.0 m. La figure 1.7 représente la valeur absolue de l'écart entre la profondeur estimée et la vraie profondeur, moyennée sur les 144 fenêtres et exprimé en

pourcentage, en fonction de la largeur de la fenêtre. Ce critère est désigné comme l'écart absolu moyen à la valeur vraie.

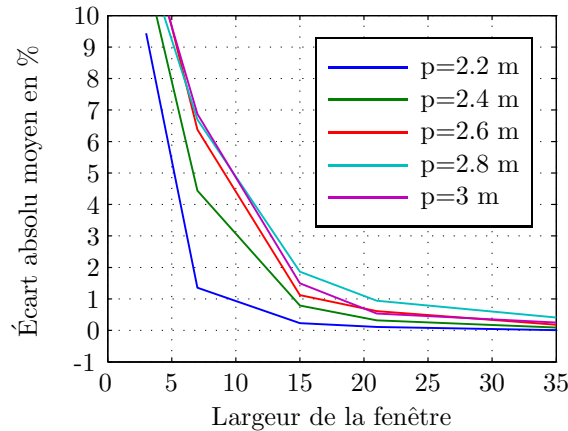


Fig. 1.7: Écart absolu moyen en % à la valeur vraie en fonction de la largeur des fenêtres traitées par l'algorithme GLDFD.

L'allure générale des courbes est semblable pour toutes les FEP, avec notamment une décroissance rapide de l'écart lorsque la largeur de la fenêtre varie entre 3 et 15 pixels. Après une vingtaine de pixels, les courbes semblent atteindre une asymptote proche de 0. Ces courbes illustrent également que les écarts les plus élevés sont obtenus pour les FEP les plus étendues (profondeur croissante). L'étude de cette tendance sera approfondie au chapitre 3. Cette simulation montre que le choix d'une fenêtre de données de taille 21×21 pixels est un bon compromis entre le temps de calcul, qui augmente avec le nombre de données, et l'écart absolu moyen de l'estimation de profondeur.

1.3.1.5 Comparaison avec le critère GCV

Le critère de validation croisée généralisé (GCV) est un critère proposé dans la littérature pour optimiser le paramètre de régularisation [Golub *et al.*, 1979; Wahba, 1985; Reeves et Mersereau, 1992; Neumaier, 1998]. Dans [Reeves et Mersereau, 1992], il est notamment utilisé pour estimer les paramètres du flou et de la scène dans le cas d'un modèle ARMA. Avec notre formalisme, il est possible d'écrire ce critère de la manière suivante :

$$\text{GCV}(\psi) = \frac{N \|P_\psi \mathbf{y}\|^2}{(\text{tr} P_\psi)^2} \quad (1.52)$$

où tr désigne la trace d'une matrice. En utilisant la décomposition GSVD et les notations définies pour le critère GL, le critère GCV s'écrit :

$$\text{GCV} = \left(\sum_{i=1}^N w_i \right)^{-2} \left(N \sum_{i=1}^N (w_i z_i)^2 \right), \quad (1.53)$$

L'objectif de la simulation suivante est de comparer les performances du critère GL et du critère GCV. La figure 1.8 montre les résultats d'estimation de profondeur des deux critères pour une même série de fenêtres générées de sorte que $\alpha = 10^{-3}$.

La figure 1.8 montre que le critère GCV présente plusieurs valeurs de profondeur aberrantes, ce qui n'est pas le cas du critère GL.

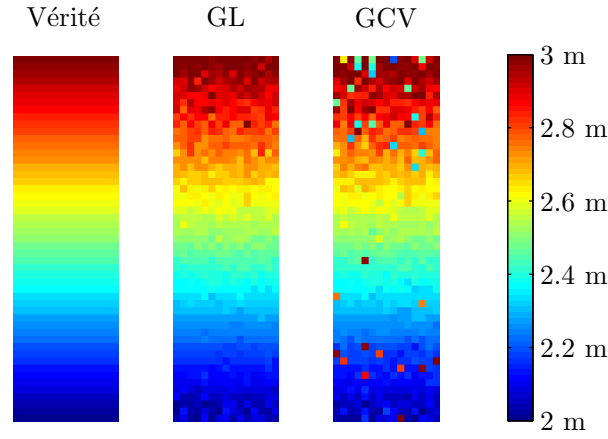


Fig. 1.8: Comparaison des performances d'estimation de profondeur des critères GCV et GL sur les mêmes données dans le cas de scènes vérifiant l'a priori gaussien.

1.3.1.6 Absence de la vraie FEP dans la famille de FEP potentielles

Dans les simulations précédentes, la FEP utilisée pour générer les images simulées fait partie de la famille de FEP potentielles utilisée par l'algorithme GLDFD. L'objectif de cette simulation est d'observer le comportement de l'algorithme lorsque cette FEP ne fait pas partie de la famille. Ici, les images sont générées avec des FEP associées à la famille de FEP simulées pour des profondeurs variant entre 2.0 m et 2.8 m par pas de 2 cm. Cette famille de FEP est notée F. Nous construisons également une famille de FEP, notée F', qui contient les FEP associées aux profondeurs comprises entre 2.1 m et 2.9 m par pas de 2 cm. La figure 1.9 présente les résultats d'estimation de profondeur autour de 2.3 m, dans le cas où l'algorithme utilise respectivement la famille F et la famille F'.

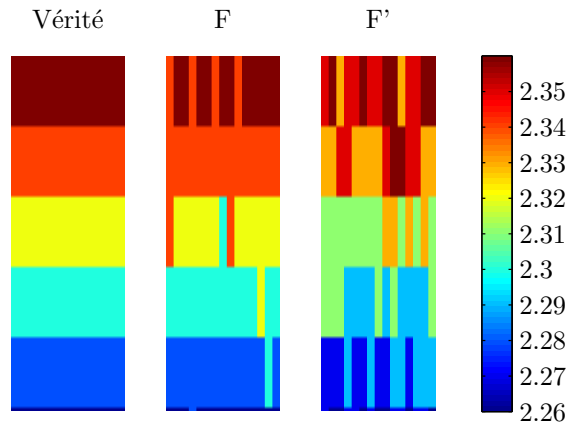


Fig. 1.9: Résultats d'estimation de profondeur, en utilisant la famille de FEP ayant servi à simuler les images (famille F) ou la famille décalée (famille F').

Les résultats d'estimation de FEP montrent que pour chaque vraie profondeur, les valeurs de profondeurs estimées encadrent la vraie profondeur. Ainsi, en l'absence de la vraie FEP dans la base de FEP potentielles, l'algorithme GLDFD sélectionne les FEP les plus proches.

1.3.2 Scènes naturelles

Jusqu'à présent l'algorithme GLDFD a été utilisé sur des scènes vérifiant l'a priori gaussien utilisé pour modéliser la scène dans l'expression du critère GL. Dans ce paragraphe, les images sont simulées à partir de scènes naturelles, avec un bruit blanc gaussien d'écart-type $\sigma_b = 0.01$, sachant que l'intensité de la scène est normalisée pour être comprise entre 0 et 1. La figure 1.10 présente quatre scènes utilisées dans les simulations suivantes.

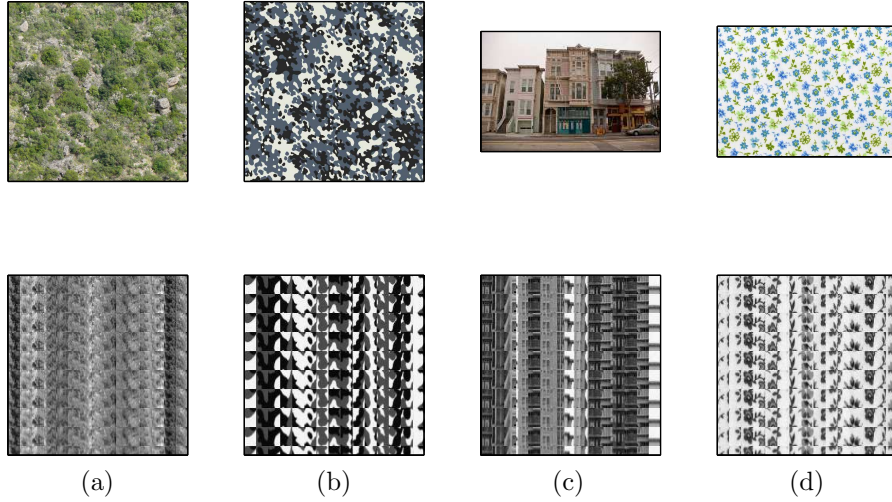


Fig. 1.10: Quatre scènes naturelles utilisées pour tester l'algorithme GLDFD.

Chaque scène (a)-(d) est successivement convoluée aux FEP de la famille de FEP potentielles. Pour chaque scène et pour chaque FEP, quinze fenêtres de taille 21×21 pixels sont extraites de l'image simulée et concaténées horizontalement. La série de fenêtres est ensuite concaténée verticalement pour les différentes FEP. Des exemples d'images ainsi obtenues pour quelques FEP sont présentées à la figure 1.10. Notons que la position des quinze fenêtres dans l'image est choisie aléatoirement mais qu'elle ne varie pas d'une FEP à l'autre. Cette opération est reproduite pour chacune des scènes de la figure 1.10. Un bruit blanc gaussien d'écart-type 0.01 est également ajouté à chaque fenêtre, sachant que les scènes sont normalisées.

1.3.2.1 Performance d'estimation de profondeur sur des scènes naturelles

La figure 1.11 présente les résultats d'estimation de profondeur obtenus avec le critère GL pour chaque scène de la figure 1.10. En comparant ces résultats avec ceux obtenus à la figure 1.4 dans le cas de scènes vérifiant l'a priori gaussien, il apparaît que la variance des résultats d'estimation de profondeur est plus importante sur les scènes naturelles, cependant les résultats d'estimation de profondeur restent cohérents avec la vérité. Pour certaines fenêtres l'algorithme renvoie toutefois des valeurs aberrantes de profondeur.

1.3.2.2 Comparaison des décompositions en Fourier et GSVD

La décomposition GSVD a été choisie pour implémenter le critère GL car, compte tenu du fait que nous considérons des convolutions de type « valide » dans lesquelles le support de l'image y est inférieur à celui de la scène x , elle permet de ne faire aucune approximation au niveau des bords de l'image. Nous justifions ici ce choix en comparant les résultats

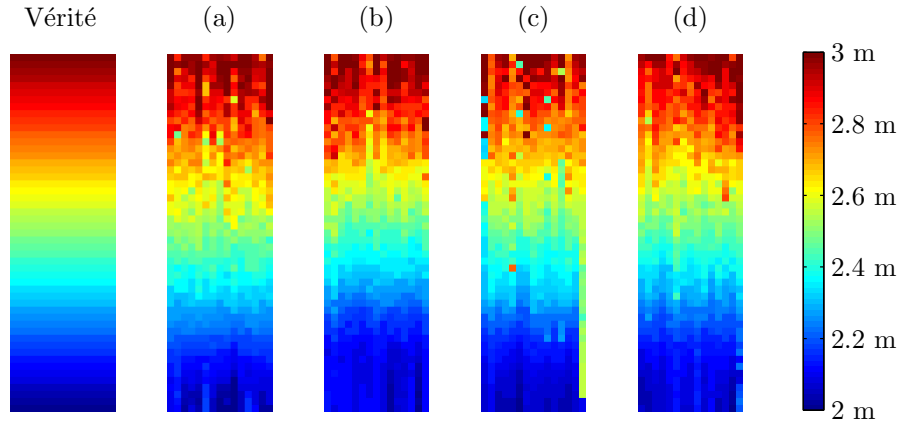


Fig. 1.11: Résultats de l'estimation de profondeur avec l'algorithme GLDFD sur des extraits des scènes (a) à (d) présentées à la figure 1.10.

d'estimation de FEP obtenus en implémentant le critère GL avec la GSVD avec ceux obtenus avec la décomposition en transformée de Fourier. Deux types de scènes sont considérées ici, une scène très haute fréquence constituée d'un bruit blanc gaussien et une scène naturelle.

Dans le premier exemple, l'image est une concaténation verticale d'une scène haute fréquence générée à partir d'un bruit blanc gaussien de moyenne nulle et d'écart-type 0.5 convolué à une FEP correspondant à des profondeurs décroissantes. Un bruit blanc gaussien est ajouté au résultat. La figure 1.12 présente les résultats d'estimation de profondeur en utilisant le critère GL calculé avec la décomposition en Fourier ou la décomposition GSVD, pour des fenêtres de tailles 21×21 pixels.

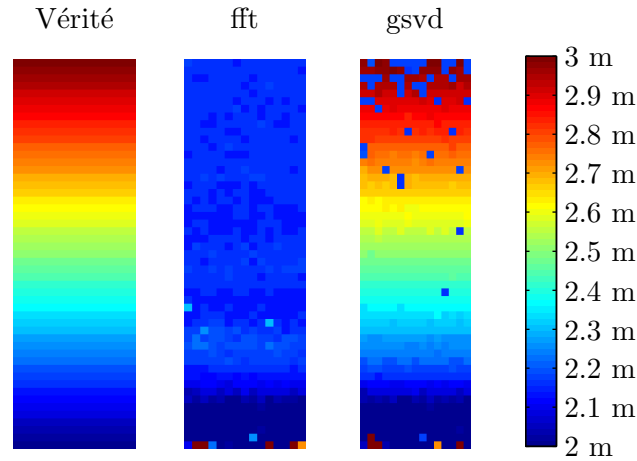


Fig. 1.12: Résultats d'estimation de profondeur à l'aide du critère GL implémenté à l'aide de la décomposition en Fourier et de la décomposition GSVD pour des fenêtres de tailles 21×21 .

Contrairement aux résultats obtenus avec la décomposition GSVD, l'estimation de profondeur avec la décomposition en Fourier donne des résultats aberrants. Nous pouvons cependant remarquer que ces résultats semblent biaisés vers une profondeur de 2.2 m. En enlevant de la famille de FEP potentiels les FEP les plus étroites, autrement dit les plus faibles profondeurs, nous pouvons constater que les résultats d'estimation de profondeur

avec la décomposition en Fourier s'améliorent. La figure 1.13 présente notamment les résultats d'estimation de profondeur en utilisant la décomposition en Fourier pour différentes tailles de fenêtre et le résultat donné par le critère GL dans le cas de fenêtres de tailles 21×21 pixels, pour une famille de FEP correspondant aux profondeurs comprises entre 2.4 m et 3 m. Pour des petites fenêtres, la décomposition en Fourier est peu satisfaisante,

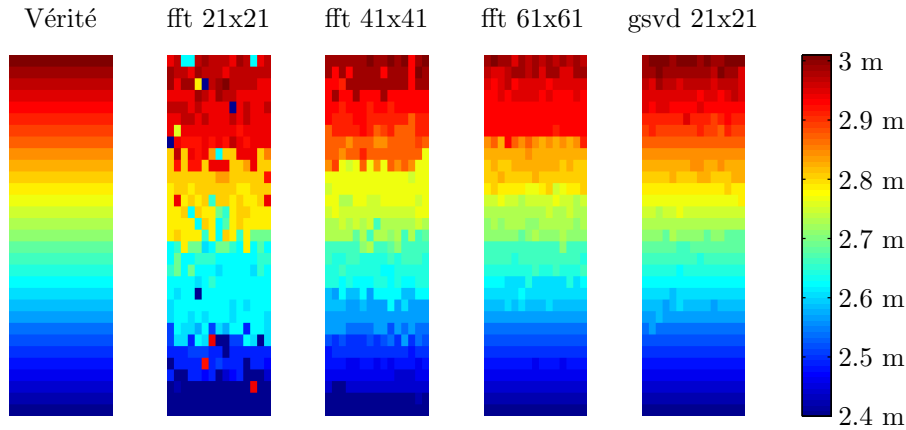


Fig. 1.13: Résultats d'estimation de profondeur à l'aide du critère GL implémenté à l'aide de la décomposition en Fourier pour différentes tailles de fenêtres et de la décomposition GSVD pour des fenêtres de tailles 21×21 , en enlevant de la famille de FEP potentielles les FEP les plus étroites (profondeurs inférieures à 2.4 m).

mais les résultats s'améliorent lorsque la taille de fenêtre augmente. Ainsi, il faut avoir une taille de fenêtre de 61×61 pour que les performances de la décomposition en Fourier soient comparables à celles de la décomposition GSVD obtenues pour une fenêtre de taille 21×21 pixels. Ensuite, le critère GL implémenté en utilisant la décomposition de Fourier est testé sur les mêmes images que celles utilisées à la figure 1.11. La figure 1.14 présente les résultats d'estimation de profondeur obtenus. Dans ce cas, tous les résultats d'estimation

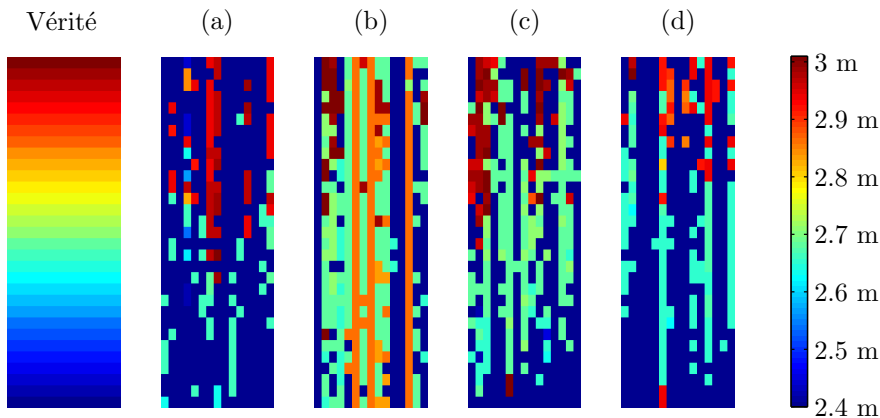


Fig. 1.14: Résultat d'estimation de profondeur avec le critère GL et la décomposition de Fourier sur les scènes (a) à (d) de la figure 1.10.

de profondeur sont aberrants.

Ces deux exemples montrent que la décomposition en Fourier fonctionne lorsque la scène contient beaucoup de hautes fréquences, que la taille de la fenêtre n'est pas trop petite et

pour des FEP de taille importante. Dans ce cas, les effets de bords liés à l'approximation circulante sont moins importants. Par contre, lorsque la scène est naturelle, la décomposition en Fourier donne des résultats aberrants, quelle que soit la taille de la fenêtre traitée. Notons que d'autres méthodes pourraient être envisagées, une première solution consisterait à appliquer une apodisation sur la fenêtre de l'image. Une autre solution consisterait à utiliser une hypothèse de scène « en miroir » et une décomposition en cosinus (DCT). Cette méthode est moins sensible aux effets de bords que la transformée de Fourier, mais elle implique une hypothèse de flou symétrique. La solution proposée par GSVD nous a ainsi semblé plus simple et permet d'exploiter directement toutes les données sans faire d'hypothèse sur la forme de la FEP ou sur les contours de l'image.

1.3.3 Bilan

Dans cette section, nous avons testé l'algorithme GLDFD sur des images simulées. Dans un premier temps, nous avons construit des images à partir de scènes vérifiant l'a priori gaussien et nous avons montré que l'algorithme GLDFD permettait d'estimer correctement les deux paramètres dont il dépend, à savoir la FEP et le rapport signal à bruit. De plus, le critère GL s'est avéré plus performant que le critère GCV sur les mêmes données. Enfin, l'absence de la vraie FEP dans la famille de FEP se traduit par la sélection des FEP les plus proches. Nous avons ensuite testé l'algorithme GLDFD proposé sur des images naturelles. Des valeurs aberrantes apparaissent parfois, car la scène ne vérifie plus exactement l'a priori utilisé dans l'algorithme, mais dans l'ensemble les résultats obtenus sont proches de ceux obtenus avec des scènes vérifiant l'a priori gaussien. Nous avons également montré que la décomposition en Fourier des matrices H et D pour calculer le critère GL ne fonctionne pas sur des scènes naturelles ou des fenêtres de petites tailles contrairement à la décomposition en GSVD que nous proposons. Dans la section suivante, l'algorithme GLDFD est comparé à trois algorithmes de la littérature développés pour des applications d'estimation de flou de bougé ou de flou de défocalisation.

1.4 Comparaison avec plusieurs méthodes de la littérature

Dans cette section, l'algorithme GLDFD est comparé à plusieurs méthodes de la littérature dans les cas du flou de bougé, du flou de défocalisation modélisé par une gaussienne et du flou de défocalisation généré par un imageur à pupille codée. Les différentes méthodes de la littérature comparées ici utilisent des techniques de régularisation des résultats pour améliorer les cartes de FEP. Ces techniques donnent des cartes de profondeur très propres mais sont souvent coûteuses en temps de calcul et dépendent de nombreux paramètres. Dans ce chapitre, notre objectif est de fournir rapidement une estimation de profondeur dite « brute » dans l'objectif du temps réel. Nous comparons donc ici, dans la mesure du possible, notre méthode avec celles proposées dans la littérature avant l'étape de régularisation.

1.4.1 Flou de bougé

La méthode de [Chakrabarti *et al.*, 2010] a été décrite dans la section 1.1.3.5. Les codes relatifs à cette méthode sont disponibles sur le site Internet ¹. À partir des codes fournis sur Internet, il est possible de comparer les performances d'estimation locale de FEP à partir du critère de vraisemblance de Chakrabarti ou du critère GL. Nous effectuons cette

¹<http://www.eecs.harvard.edu/~ayanc/svblur/code/>.

comparaison ici, sur des images simulées et réelles. La famille de FEP potentielles considérée est constituée de FEP 1D de largeur L variant entre 1 et 15 pixels et les fenêtres de données sont de taille 1×61 pixels.

1.4.1.1 Images simulées

Dans cette simulation, quatre scènes naturelles sont convoluées à une des FEP de la famille de FEP potentielles. Un bruit blanc gaussien d'écart-type σ_b est ajouté au résultat, sachant que toutes les scènes sont normalisées. Une fenêtre de taille 25×123 pixels est extraite de chaque image simulée et les FEP sont estimées sur ces fenêtres avec les deux méthodes. Comme la méthode de Chakrabarti donne des résultats sur tous les pixels, le pas dans l'algorithme GLDFD est de 1. La figure 1.15 présente des extraits des quatre scènes utilisées ainsi que la moyenne et l'écart-type des indices estimés suivant les deux critères, pour plusieurs FEP. Ces valeurs sont calculées en concaténant les résultats obtenus sur les quatre fenêtres. Deux niveaux de bruit sont considérés $\sigma_b = 0.01$ et $\sigma_b = 0.001$.

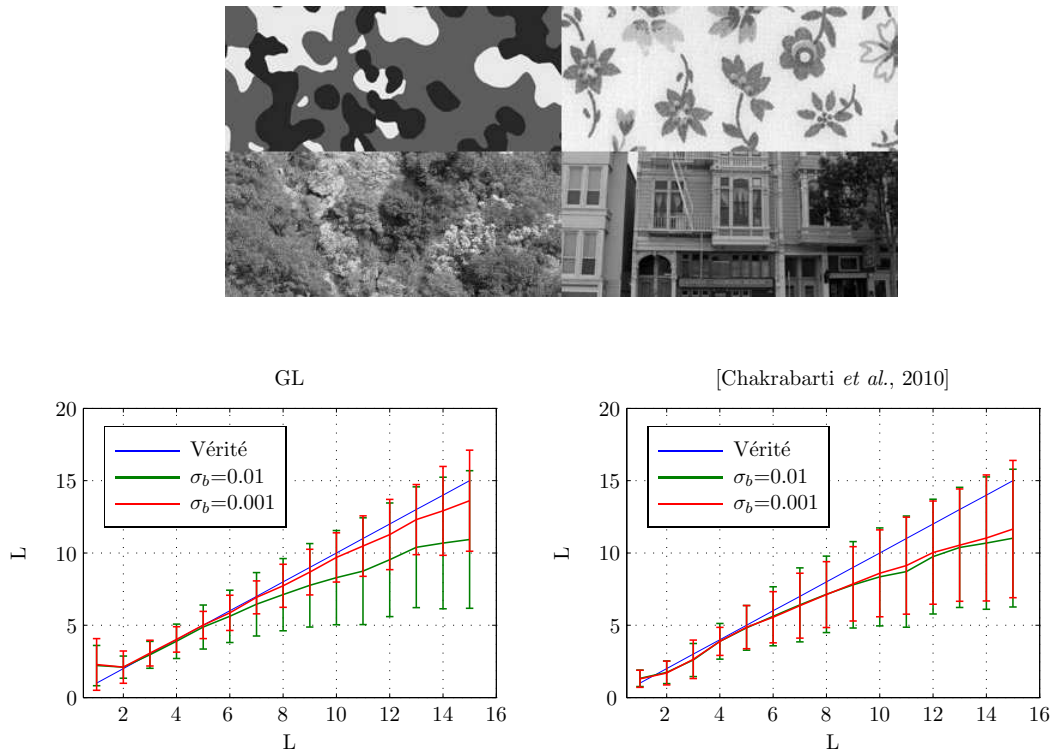


Fig. 1.15: Résultats d'estimation de FEP obtenus avec le critère GL et la méthode de [Chakrabarti *et al.*, 2010] moyennés à partir d'images simulées pour différentes valeurs de bruit.

Pour un niveau de bruit faible, les résultats obtenus avec l'algorithme GLDFD présente de meilleurs résultats d'estimation de FEP, en termes de biais comme de variance des résultats, sauf pour la FEP la plus étroite, où notre algorithme sélectionne la largeur 2. Pour un bruit plus élevé, les deux méthodes donnent des résultats équivalents.

1.4.1.2 Images réelles

Les figures 1.16 et 1.17 montrent les cartes d'indice de FEP les plus probables, obtenues avec le critère de Chakrabarti et le critère GL, sur deux des images réelles fournies par Chakrabarti. Dans les résultats obtenus avec le critère GL, le recouvrement des fenêtres est de 30%.

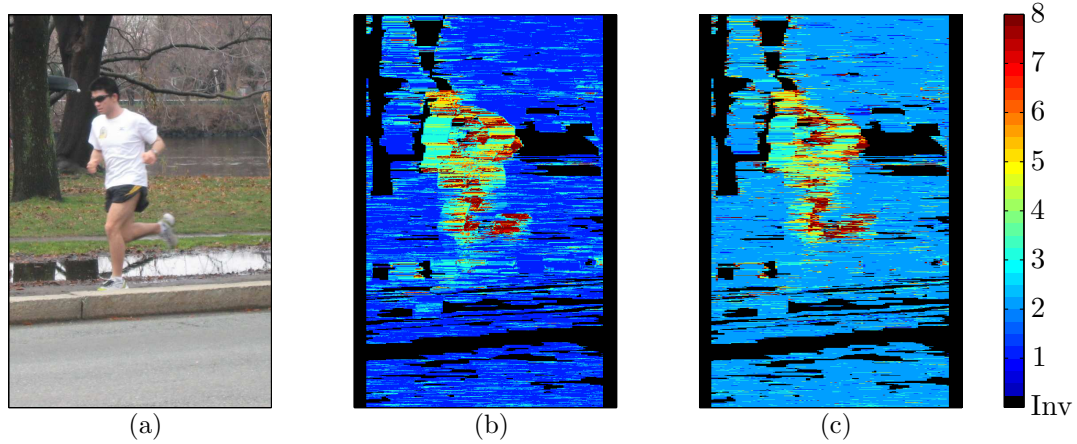


Fig. 1.16: (a) Image réelle associée à la référence [Chakrabarti *et al.*, 2010]. Résultats d'estimation de FEP avec : (b) la méthode [Chakrabarti *et al.*, 2010], (c) l'algorithme GLDFD. Les régions en noir correspondent aux régions non texturées rejetées par le filtre de Canny.

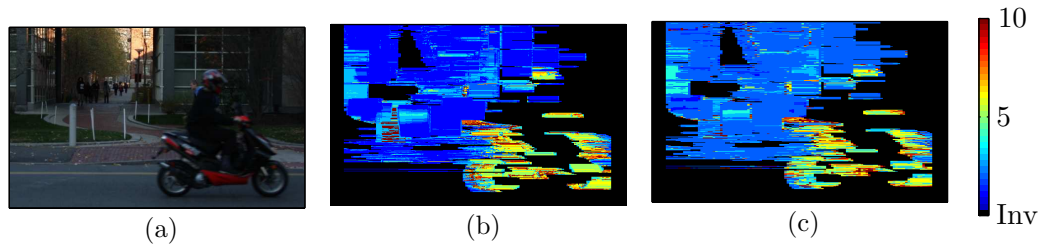


Fig. 1.17: (a) Image réelle associée à la référence [Chakrabarti *et al.*, 2010]. (b) Résultats d'estimation de FEP avec : (b) de la méthode [Chakrabarti *et al.*, 2010], (c) l'algorithme GLDFD. Les régions en noir correspondent aux régions non texturées rejetées par le filtre de Canny.

Les deux méthodes permettent d'identifier l'objet mobile de l'image et donnent des résultats comparables. Toutefois le fond pour le critère GL est associé à la FEP de largeur 2 au lieu de 1. Ce phénomène était déjà visible en simulation.

Si les résultats d'estimation de FEP à une dimension sont comparables en utilisant les deux critères sur des images réelles, la méthode proposée par Chakrabarti ne permet pas de traiter le cas de plusieurs objets mobiles dans des directions différentes. En effet, la première étape de son algorithme consiste à estimer dans toute l'image la direction et la taille du flou le plus probable sous l'hypothèse d'un seul objet mobile dans l'image. Or, l'algorithme GLDFD permet de traiter le cas de plusieurs objets mobiles dans des directions différentes. En effet, puisqu'il fonctionne également avec des FEP à deux dimensions, il est

possible de construire une famille de FEP constituée de différentes directions potentielles de déplacement. L'algorithme estime alors localement le flou de bougé le plus probable. La figure 1.18 est un exemple de résultats d'estimation de FEP obtenus dans ce cas. L'image présentée en (a) est composée de deux objets mobiles de direction et de vitesse différentes. La famille de FEP potentielles est composée de FEP à deux dimensions de taille 15×15 pixels dont les valeurs non nulles sont des segments horizontaux ou verticaux de longueur variable. La figure (b) montre la carte des objets mobiles obtenue avec le critère GL. La couleur verte est associée aux déplacements verticaux et la couleur bleue aux déplacements horizontaux. L'orientation du mouvement des deux objets mobiles est correctement identifiée et l'intensité différente des deux objets est cohérente avec leur différence de vitesse.

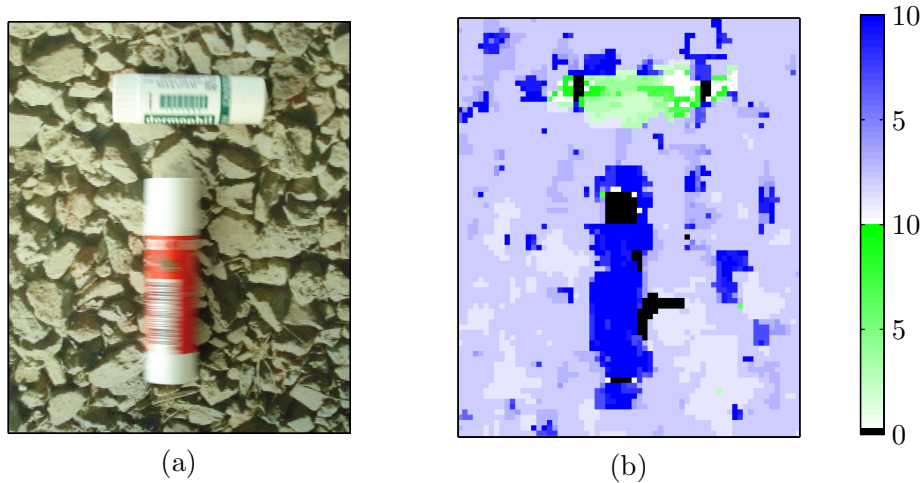


Fig. 1.18: Résultat d'estimation du flou de bougé obtenu avec l'algorithme GLDFD sur une image réelle avec deux objets mobiles dans des directions différentes.

1.4.1.3 Conclusion

En conclusion, la comparaison des deux algorithmes sur des images simulées a montré que dans le cas d'un bruit faible, la variance des résultats de l'algorithme GLDFD est plus faible que celle des résultats obtenus avec l'algorithme de Chakrabarti. La comparaison des deux algorithmes sur des images réelles a montré que les deux méthodes présentent des résultats comparables. En revanche, l'algorithme de Chakrabarti est restreint au cas d'un seul objet mobile et il est développé uniquement dans le cas de flou de bougé, contrairement à l'algorithme GLDFD.

1.4.2 Estimation du flou de défocalisation dans le cas d'une optique classique

Nous avons vu dans l'état de l'art que la FEP est souvent modélisée par une gaussienne et plusieurs articles exploitent cette information en estimant la profondeur au niveau des gradients images directionnels, en modélisant la scène par une fonction de Heaviside [Tai et Brown, 2009; Hu et De Haan, 2007; Tai *et al.*, 2009; Torrea et Fernandes, 2005; Zhuo et Sim, 2011; Zhuo et Sim, 2009; Cao *et al.*, 2011]. L'algorithme GLDFD est comparé ici à l'algorithme présenté dans [Zhuo et Sim, 2011]. La famille de FEP est constituée de gaussiennes d'écart-type σ_g variant entre 0.5 et 4 par pas de 0.1, définis sur une fenêtre de

taille 21×21 pixels. La méthode de Zhuo n'étant pas disponible publiquement, elle a été recodée selon les informations fournies dans les articles [Zhuo et Sim, 2009; Zhuo et Sim, 2011]. L'image est convoluée par une gaussienne d'écart-type égal à $\sigma_0 = 1$. Puis le rapport entre le module des gradients de l'image reconstruite et de l'image de départ est calculé au niveau des contours détectés par la méthode Canny, ce qui permet d'estimer l'écart-type du flou initial par la formule (1.20). Ce procédé est sensible au bruit et à l'interférence entre les flous des contours voisins, c'est pourquoi Zhuo applique ensuite un filtre bilatéral de largeur 20 pixels. Les paramètres du filtre utilisés dans les résultats suivants sont les mêmes que ceux indiqués dans [Zhuo et Sim, 2011], à savoir un écart-type de 10% de l'intensité maximale pour la gaussienne du filtre bilatéral régularisant la carte de FEP en fonction de l'intensité des pixels et un écart-type de 10% de la taille de l'image pour la gaussienne spatiale. Pour le calcul du critère GL, les fenêtres des données sont de taille 21×21 pixels.

1.4.2.1 Images simulées

Notre méthode et celle de Zhuo sont tout d'abord testées sur les deux scènes présentées à la figure 1.19. La première scène présente de fortes discontinuités et la deuxième scène est une scène naturelle. Ces deux scènes sont converties en niveaux de gris et normalisées puis sont convoluées à une des FEP de la famille de FEP gaussiennes. Un bruit blanc gaussien d'écart-type $\sigma_b = 0.001$ ou $\sigma_b = 0.01$ est alors ajouté au résultat. Pour chaque valeur de flou gaussien et pour chaque méthode, la moyenne et l'écart-type des résultats d'estimation de FEP sont calculés et présentés à la figure 1.20.



Fig. 1.19: Deux scènes utilisées dans la comparaison entre la méthode [Zhuo et Sim, 2011] et l'algorithme GLDFD.

Pour la scène (a) présentant de nombreuses discontinuités et le plus faible niveau de bruit, les deux méthodes donnent des résultats satisfaisants avec un biais et un écart-type faibles pour des valeurs d'écart-type σ_g supérieure à 1.5. Lorsque le bruit augmente, les résultats des deux méthodes se dégradent et notamment pour le critère GL un biais constant apparaît. Pour la scène naturelle (b) et les deux niveaux de bruit, les performances de l'algorithme de Zhuo s'effondrent pour des valeurs de flou importantes. En effet, sur une scène naturelle, la modélisation du contour de la scène par une fonction de Heaviside est moins appropriée que pour la scène (a) et de plus, comme annoncé par les auteurs, il y a des interactions entre les flous des différents contours qui perturbent l'estimation. Par contre l'algorithme GLDFD permet d'estimer correctement la profondeur et notamment pour des flous importants. Notre méthode est donc plus adaptée au cas de scènes naturelles et aux flous importants. En contrepartie, un biais apparaît pour un niveau de bruit sur les scènes dont la texture est trop éloignée de l'a priori gaussien isotrope.

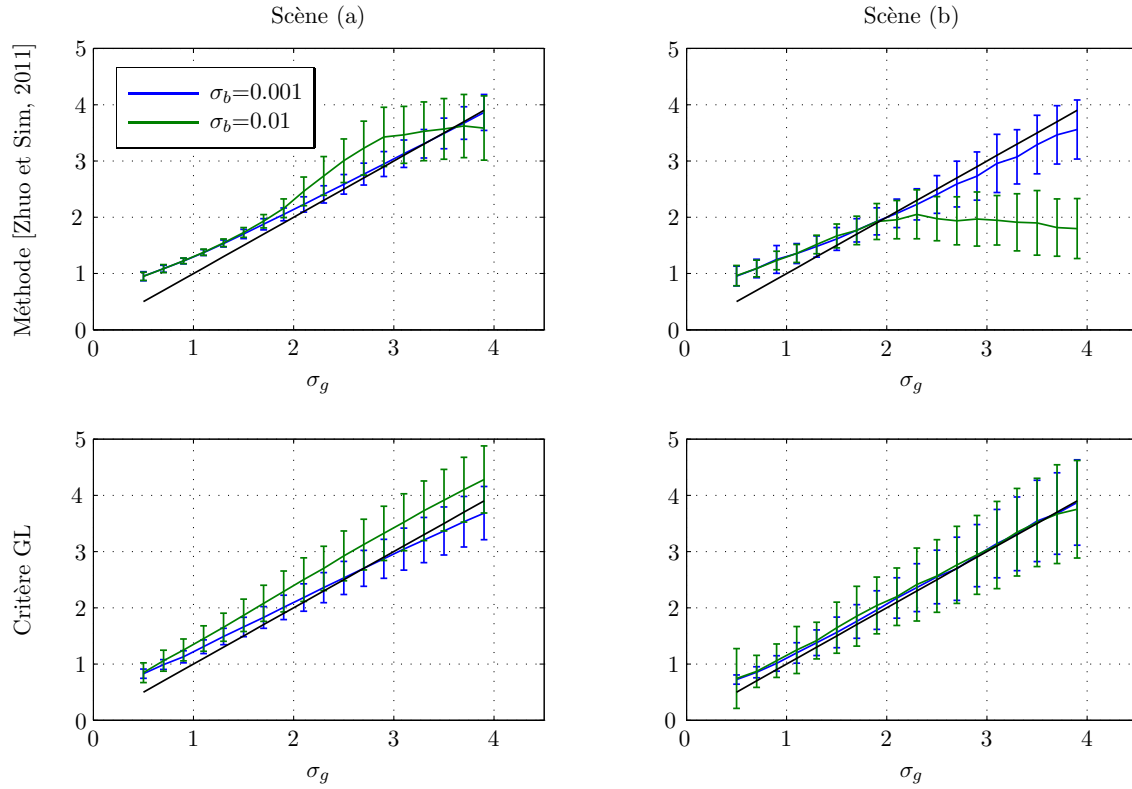


Fig. 1.20: Résultats d'estimation de FEP obtenus en utilisant la méthode [Zhuo et Sim, 2011] et l'algorithme GLDFD sur les deux scènes présentées à la figure 1.19.

1.4.2.2 Images réelles

La figure 1.21 présente les résultats d'estimation de profondeur en utilisant la méthode proposée dans ce chapitre et la méthode de Zhuo, sur deux images réelles fournies par Zhuo. Le recouvrement des fenêtres dans notre méthode est de 50%. Les deux cartes de profondeur de la figure 1.21 permettent d'identifier correctement les différents niveaux de profondeur de l'image. Le résultat obtenu avec la méthode de Zhuo ne donne des informations de profondeur que sur les gradients de l'image, ce qui impose une régularisation a posteriori. Cette régularisation augmente notablement le coût de calcul. Par contre, le résultat brut de notre méthode donne une information de profondeur dans toutes les régions texturées de l'image. Ceci illustre l'intérêt d'utiliser une fenêtre de l'image pour estimer la profondeur et non uniquement les contours.

1.4.2.3 Conclusion

En conclusion, les tests menés en simulation avec une version recodée de la méthode de Zhuo montre que cette méthode ne donne des résultats satisfaisants que sur des images à fortes discontinuités et avec un faible niveau de bruit. L'algorithme GLDFD est plus robuste au bruit notamment pour des scènes naturelles. La comparaison sur un exemple de Zhuo montre que les cartes de profondeurs obtenues avec les deux algorithmes sur une image réelle sont assez proches, cependant la méthode de Zhuo donne des résultats uniquement sur les contours, ce qui rend indispensable une régularisation pour avoir un résultat sur

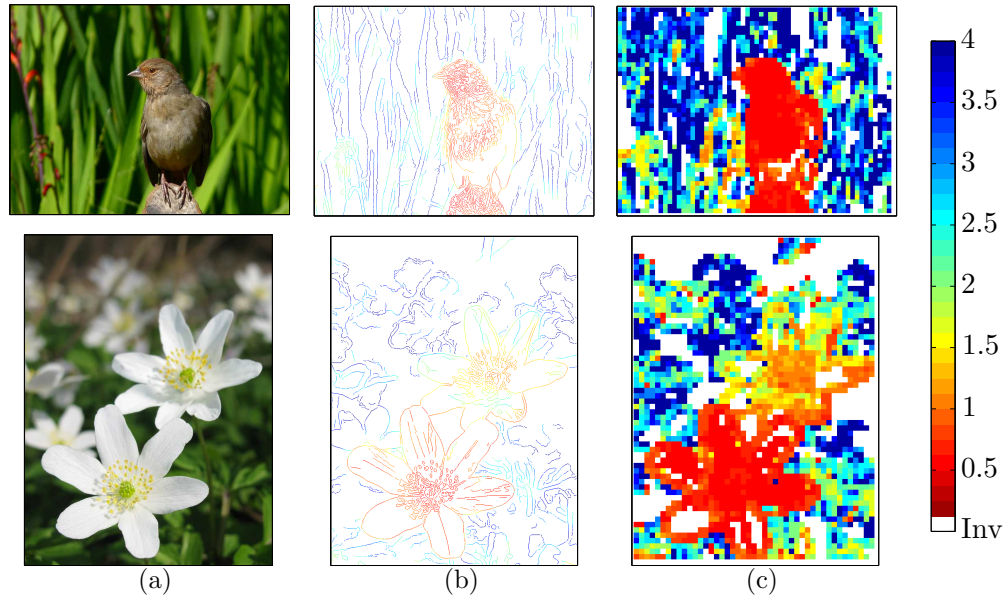


Fig. 1.21: (a) Deux images réelles issues de la base d’images de [Zhuo et Sim, 2011]. Résultats obtenu en utilisant : (b) la méthode [Zhuo et Sim, 2011], (c) le critère GL. Les régions en blanc correspondent aux régions homogènes rejetées par le filtre de Canny.

l’image entière. De notre côté, nous fournissons une carte plus dense qui permet d’identifier les niveaux de profondeur de l’ensemble de la scène. Enfin, la méthode de Zhuo est dédiée au cas d’une pupille circulaire classique, ce qui n’est pas le cas de l’algorithme que nous proposons.

1.4.3 Flou de défocalisation dans le cas d’une optique à pupille codée

Dans cette section, sont considérés deux articles de la littérature proposant chacun une méthode d’estimation de profondeur dans le cas d’un imageur à pupille codée. La première méthode correspond à la référence [Martinello et Favaro, 2011b] qui a été présentée à la section 1.1.3.5. Cette méthode n’étant pas publiquement disponible, elle a été reprogrammée au DTIM par l’ingénieur V. Leung. Les résultats de cette méthode dépendent de la dimension du sous-espace sur lequel les données sont projetées. Martinello ne donnant pas d’information sur comment choisir cette valeur, elle est constituée un paramètre de notre implémentation, noté n_λ . Les résultats produits par cette méthode sont présentés ici pour différentes valeurs de ce paramètre. Les comparaisons des performances d’estimation de profondeur de l’algorithme GLDFD et de la méthode de Martinello sont conduites sur des images simulées.

Les images réelles utilisées dans [Martinello et Favaro, 2011b] n’étant pas disponibles publiquement, l’algorithme GLDFD est ensuite testé sur des images réelles à partir des données fournies par [Levin *et al.*, 2007b] sur le site Internet ² et comparé aux résultats de Levin avant régularisation par un *graphcut*.

²<http://groups.csail.mit.edu/graphics/CodedAperture/>

1.4.3.1 Images simulées

Les FEP de défocalisation d'un imageur idéal à pupille codée sont générées avec la méthode décrite dans l'annexe B. Les caractéristiques de cet imageur sont données dans le tableau 1.2. La pupille de cet imageur est de la forme proposée dans [Levin *et al.*, 2007b].

Focale	35 mm
Ouverture	$F/\# = 2.8$
p_x	$6 \mu m$
p_0	1.9 m

Tableau 1.2: Caractéristiques du système optique idéal utilisé dans les simulations d'un imageur à pupille codée.

Les FEP sont générées pour des distances comprises entre 1.9 m et 3.9 m avec un pas de 5 cm. La figure 1.22 montre des exemples de FEP simulées pour plusieurs profondeurs.

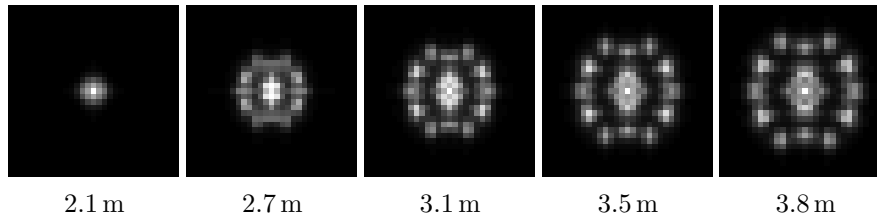


Fig. 1.22: Exemples de FEP simulées en utilisant la pupille codée de [Levin *et al.*, 2007b].

Dans la simulation suivante, 12 scènes naturelles sont convoluées à une des FEP de la famille. Un bruit blanc gaussien d'écart-type σ_b est ajouté au résultat. Cinq fenêtres sont extraites aléatoirement de chacune des 12 images obtenues. Ces fenêtres sont concaténées pour être traitées une à une par la méthode Martinello et par notre méthode. Cette opération est répétée pour toutes les FEP de la famille de FEP potentielles. La figure 1.23(a) présente les résultats d'estimation de profondeur sans bruit de la méthode de Martinello pour différentes valeurs de n_λ et de notre méthode. La taille des fenêtres est de 21×21 pixels.

En l'absence de bruit, le nombre de valeurs aberrantes obtenues avec la méthode de Martinello est plus important qu'avec l'algorithme GLDFD. Le nombre de valeurs aberrantes augmente également avec la dimension de n_λ . La simulation précédente est reproduite avec un écart-type de bruit de 0.01 et les nouveaux résultats obtenus sont présentés à la figure 1.24. Pour un bruit plus élevé, la méthode de Martinello n'estime pas la profondeur correctement, quelle que soit la valeur de N_λ . Par contre l'algorithme GLDFD donne des résultats d'estimation corrects bien qu'avec une variance plus importante qu'en l'absence de bruit. Ainsi, dans cette simulation le critère GL s'avère être plus robuste au bruit que la méthode de Martinello. L'algorithme de Martinello ayant été reproduit, la base d'apprentissage que nous avons utilisée est différente de celle utilisée dans l'article [Martinello et Favaro, 2011b], ce qui peut expliquer que nos résultats soient moins robustes au bruit que ceux présentés par Martinello. D'un autre côté, cela illustre la difficulté à reproduire une méthode basée sur un apprentissage, ce qui motive le développement de méthode non supervisée.

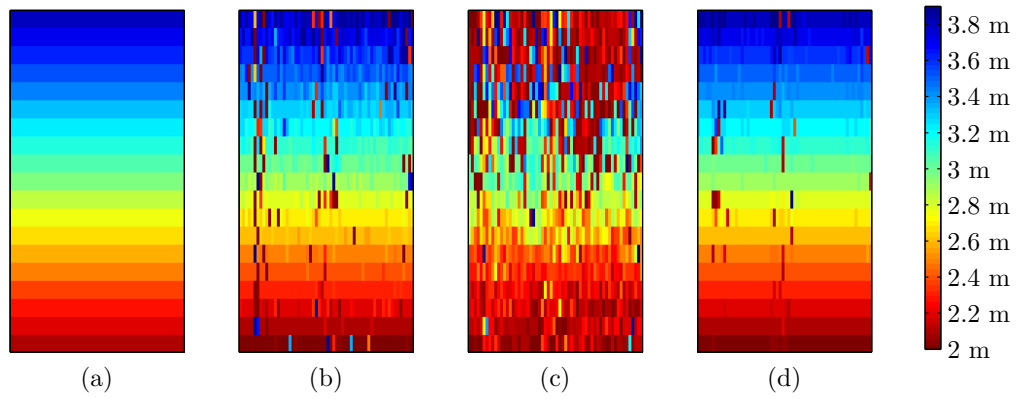


Fig. 1.23: Comparaison des résultats d'estimation de FEP obtenus avec la méthode de Martinello et l'algorithme GLDFD sans bruit sur des scènes naturelles. (a) Vérité. (b) et (c) Résultats de [Martinello et Favaro, 2011b] avec respectivement $n_\lambda = 200$ et $n_\lambda = 400$. (d) Résultats de l'algorithme GLDFD.

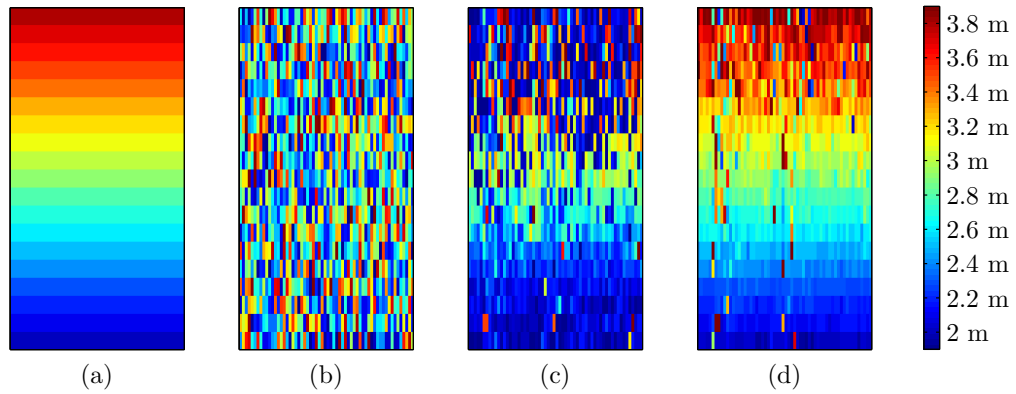


Fig. 1.24: Résultats d'estimation de FEP obtenus avec la méthode de Martinello avec le critère GL avec un bruit d'écart-type 0.01 sur des scènes naturelles. (a) Vérité. (b) et (c) Résultats de [Martinello et Favaro, 2011b] avec respectivement $n_\lambda = 200$ et $n_\lambda = 400$. (d) Résultats du critère GL.

1.4.3.2 Images réelles

Levin et ses collaborateurs fournissent, les données qui ont servi à produire leur résultats [Levin *et al.*, 2007b]. Ainsi, les FEP de l'imageur à pupille codées, calibrées pour 9 profondeurs comprises entre 2 m et 3 m, sont publiquement disponibles ainsi que plusieurs acquisitions. La figure 1.25 présente les résultats d'estimation de profondeur obtenus avec l'algorithme GLDFD sur certaines de ces données. Ces résultats sont comparés à ceux obtenus par Levin avant l'étape de régularisation. Notons que nous inversons ici l'échelle de couleur représentant les profondeurs, afin d'avoir des résultats comparables à ceux de Levin.

Dans les régions texturées de l'image, l'estimation de profondeur avec le critère GL est cohérente avec les résultats de [Levin *et al.*, 2007b] mais avec une variance supérieure. La technique de Levin donne des résultats dans les régions homogènes de l'image, même avant régularisation. Ceci peut s'expliquer par le fait que l'étape d'estimation de profondeur de

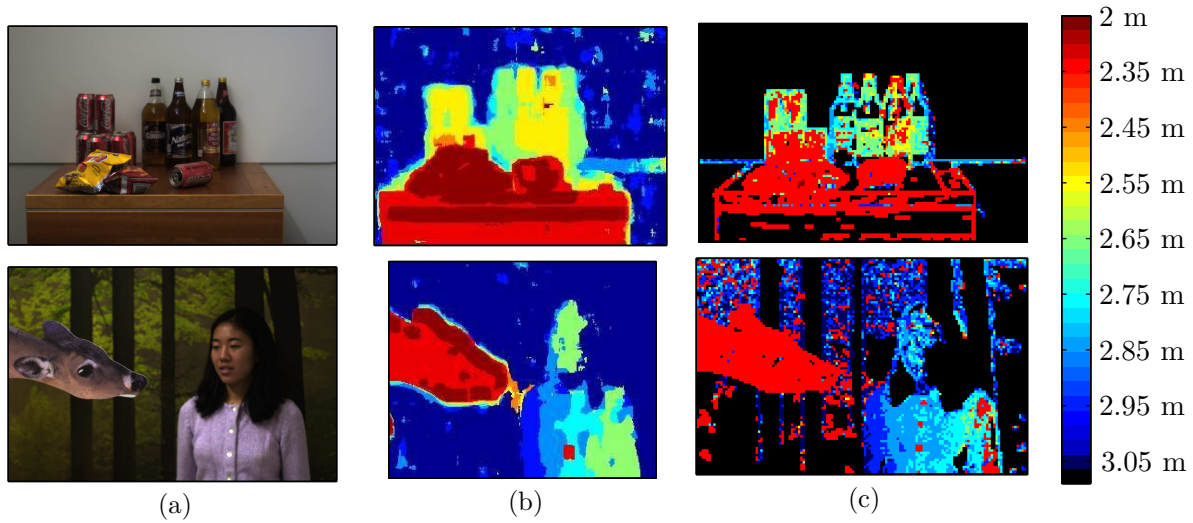


Fig. 1.25: (a) Images issues de [Levin *et al.*, 2007b]. (b) Résultats de l'estimation de profondeur par la méthode de [Levin *et al.*, 2007b] avant l'étape de régularisation. (c) Résultats obtenus avec l'algorithme GLDFD. Les régions en noir correspondent aux régions homogènes rejetées par le filtre de Canny.

Levin fait appel à une déconvolution de l'image par toutes les FEP de la famille et cette déconvolution est régularisée en utilisant un a priori qui favorise les régions constantes par morceaux.

1.4.3.3 Conclusion

En conclusion, l'algorithme que nous proposons est plus précis et plus robuste au bruit que la méthode qui a été implémentée à partir de [Martinello et Favaro, 2011b] et donne des résultats comparables à ceux de [Levin *et al.*, 2007b], cependant, contrairement aux méthodes de Levin et de Martinello, l'algorithme GLDFD n'est pas supervisé, ce qui le rend plus facilement reproductible.

1.4.4 Bilan des comparaisons menées

La comparaison de l'algorithme GLDFD avec trois algorithmes de la littérature illustre tout d'abord la génécité de notre algorithme qui fonctionne pour les trois applications considérées dans ces références, à savoir l'estimation du flou de bougé et l'estimation du flou de défocalisation à partir d'une optique classique ou d'une optique à pupille codée. Nos simulations montrent que, le critère GL s'avère plus robuste au bruit que les méthodes déduites du travail de Zhuo et de Martinello. Enfin, notre algorithme n'est pas supervisé, ce qui le rend plus facilement reproductible et indépendant d'une base d'apprentissage.

1.5 Discussions

1.5.1 Temps de calcul

La question du temps de calcul des cartes de profondeur n'a pas été abordée jusqu'à présent. Le temps nécessaire à la minimisation du critère GL sous `Matlab` pour une fenêtre de

taille 21×21 pixels est de 7 ms, avec un processeur CPU Intel Core 2 Duo de fréquence 3GHz. Le temps nécessaire à la production d'une carte de profondeur dépend également de la taille de l'image, du recouvrement choisi et du nombre de FEP à tester. Par exemple, l'estimation de l'une des cartes de profondeur présentées au paragraphe 1.4.3.2 est de l'ordre d'une minute, sachant que l'image fait 1107×1694 pixels, qu'il y a 9 FEP et que le recouvrement est de 50%. Cependant, le calcul du critère GL pour chaque FEP et pour chaque fenêtre est parallélisable. Une implémentation parallèle sous GPU de l'algorithme GLDFD est en cours au DTIM. Les premiers résultats obtenus sont dans la discussion finale de ce manuscrit (chapitre 5).

1.5.2 Hypothèse de distribution isotrope

L'a priori gaussien utilisé dans ce chapitre pour modéliser la scène suppose que les gradients ont une distribution isotrope. En réalité de nombreuses régions d'une image présentent des valeurs de gradients plus importantes dans une direction particulière. Cela peut biaiser l'estimation, comme nous l'avons remarqué à la figure 1.20(a). Une solution à ce problème, qui ne sera pas développée dans cette thèse, serait de considérer un a priori gaussien ayant une variance différente selon les directions horizontales et verticales. La différence de variance pourrait être estimée localement sur l'image en calculant la direction privilégiée des gradients. Cette solution implique de calculer à l'avance les GSVD des matrices H et D pour chaque direction privilégiée ce qui augmente le nombre de variables à stocker en prétraitement.

1.5.3 Modélisation du schéma Bayer

De nos jours, les acquisitions sont très couramment réalisées avec des détecteurs couleurs dont il existe deux familles principales. Il existe tout d'abord les détecteurs dits « tri-CCD » qui utilisent par exemple un prisme pour séparer la lumière dans trois bandes spectrales et génèrent ainsi plusieurs images de même résolution grâce à plusieurs matrices de détecteurs CCD. Les autres détecteurs couleurs sont composés d'une matrice de pixels classique mais chaque pixel possède un filtre de couleurs rouge, verte ou bleue, organisé typiquement suivant un motif dit schéma Bayer présenté à la figure 1.26.

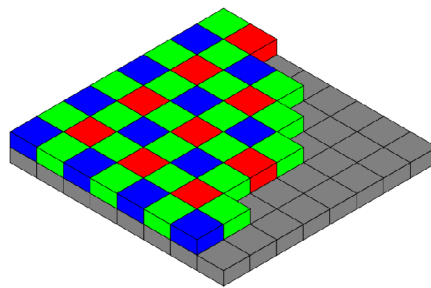


Fig. 1.26: Exemple de schéma Bayer situé sur la matrice d'un détecteur couleur.
Illustration issue de http://it.wikipedia.org/wiki/Schema_Bayer.

Dans la modélisation proposée dans ce chapitre, les données sont modélisées comme si elles provenaient d'un détecteur noir et blanc ou d'un détecteur tri-CCD. Or dans le cas d'un détecteur de type Bayer, les pixels utilisés proviennent d'un sous échantillonnage de la matrice de détecteur. Ainsi, chaque acquisition est constituée de trois sous-images correspondant à un sous-échantillonnage de l'image dans trois spectres différents. Les algorithmes

d'estimation de profondeur sont alors appliqués à l'un des canaux. Avec le formalisme utilisé dans ce chapitre, il est possible de modéliser l'effet du sous-échantillonnage lié au filtre Bayer, en enlevant les lignes de la matrice de convolution H correspondant aux pixels manquants :

$$\mathbf{y}_{bayer} = \begin{bmatrix} y_1 \\ y_3 \\ y_5 \\ \dots \end{bmatrix} = H_{bayer} \mathbf{x} = \begin{bmatrix} H(1,:) \\ H(3,:) \\ H(5,:) \\ \dots \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \end{bmatrix} \quad (1.54)$$

L'expression (1.54) suppose que la FEP échantillonnée sur tous les pixels du détecteur est connue. Cette méthode suppose donc un calibrage des FEP à l'échelle du détecteur. Dans les résultats précédents où l'algorithme GLDFD est testé sur des images réelles, l'effet du schéma Bayer est négligé et l'image traitée correspond à une des images sous-échantillonnées d'un facteur deux avec une famille de FEP potentielles également sous-échantillonnées. Cette approximation a donné des résultats satisfaisants.

1.6 Conclusion

Dans ce chapitre nous avons présenté un nouvel algorithme non supervisé d'estimation de profondeur mono-image. Cet algorithme, nommé GLDFD, prend en entrée une famille de FEP potentielles et estime sur une fenêtre de l'image la FEP la plus probable. Pour cela, nous utilisons un modèle de scène gaussien isotrope faisant intervenir une matrice de précision singulière qui rend l'a priori impropre. Il est néanmoins possible de définir une expression simple de la vraisemblance marginale des données et d'en déduire un critère de vraisemblance noté GL. Ce critère dépend de deux paramètres, à savoir le rapport signal à bruit et l'indice de la FEP. Ces deux paramètres sont estimés de manière non supervisée à l'aide d'une minimisation continue du critère GL par rapport au rapport signal à bruit et d'une minimisation discrète sur l'indice de la FEP pour chaque fenêtre de l'image. Nous avons proposé une implémentation du critère GL qui repose sur une décomposition en valeurs singulières généralisées qui permet de ne faire aucune hypothèse sur les bords de l'image ou de la scène. Cette décomposition permet de plus un traitement rapide de chaque fenêtre de l'image car elle peut être réalisée en prétraitement. Si le traitement d'une image nécessite encore un temps de calcul de l'ordre de la minute, l'algorithme que nous proposons se prête bien à la parallélisation. Enfin, le formalisme utilisé pour exprimer le critère GL est très général et nous verrons dans le chapitre 2 qu'il est facilement généralisable au cas de l'estimation de profondeur à l'aide de plusieurs images.

Nous avons testé l'algorithme GLDFD sur des images simulées avec des scènes vérifiant l'a priori gaussien ou avec des scènes naturelles. Ces tests ont montré que le critère GL permet d'estimer correctement la profondeur même sur des images naturelles qui ne suivent pas parfaitement l'a priori gaussien isotrope. Nous l'avons également testé sur des images réelles de la littérature, pour des applications d'estimation de flou de bougé et de flou de défocalisation et dans chaque cas nous avons obtenu des résultats satisfaisants d'estimation de FEP relative, l'absence de vérité terrain nous empêchant de conclure sur l'estimation de la valeur exacte de la FEP.

L'algorithme GLDFD que nous proposons est non supervisé, ce qui le rend plus facilement reproductible que les algorithmes de la littérature qui utilisent un apprentissage. De plus, nous n'imposons aucune contrainte sur la forme de la FEP dans la famille de FEP potentielles, ce qui permet de l'utiliser à la fois pour des applications d'estimation de flou de bougé comme de flou de défocalisation avec une pupille classique ou codée. Dans différents

contextes, l'utilisation du critère GL a permis d'obtenir des résultats plus performants sur des scènes réelles bruitées qu'en utilisant d'autres méthodes proposées dans la littérature.

Nous disposons donc d'un algorithme d'estimation de FEP performant qui laisse une grande liberté sur le choix de l'optique utilisée. Dans le chapitre suivant, nous nous intéressons à deux solutions optiques en particulier, l'utilisation d'une pupille codée et l'utilisation d'un imageur chromatique.

Solutions optiques favorisant l'estimation de profondeur

Dans ce chapitre nous étudions le potentiel d'estimation de profondeur de deux solutions optiques, à savoir respectivement l'utilisation d'une pupille codée et d'une optique présentant de l'aberration chromatique longitudinale.

L'utilisation d'une pupille codée à la place d'une pupille classique est étudiée dans plusieurs références de la littérature mais l'estimation précise de la profondeur n'est souvent que secondaire car l'objectif principal de ces travaux est d'obtenir une image de grande profondeur de champ. Ainsi, aucun élément précis n'est fourni permettant de s'assurer que la profondeur estimée est cohérente avec la profondeur effective. Autrement dit, aucune caractérisation de la précision d'estimation de profondeur obtenue avec un imageur muni d'une pupille codée n'a été présentée sur des images réelles. Dans ce chapitre, nous évaluons expérimentalement cette précision en combinant une optique à pupille codée et l'algorithme GLDFD décrit dans le chapitre 1. Grâce à une collaboration avec plusieurs enseignants de l'Institut d'Optique Graduate School de Palaiseau (IOGS), nous avons pu équiper un appareil photographique reflex numérique d'un objectif dont la pupille circulaire classique est remplacée par celle proposée dans [Levin *et al.*, 2007b]. Nous avons alors caractérisé la précision de ce prototype sur l'axe, en comparant les mesures de profondeur obtenues avec celles données par un télémètre. Enfin, nous proposons une comparaison visuelle des cartes de profondeur issues de notre prototype avec celles obtenues en utilisant un système actif.

La deuxième solution optique étudiée dans ce chapitre est l'utilisation d'une optique non corrigée de l'aberration chromatique longitudinale, qui produit une FEP de largeur spectralement variable à une profondeur donnée. L'optique dite « chromatique » est combinée à un détecteur couleur, qui fait l'acquisition simultanée d'images de la scène dans différentes bandes spectrales, ce qui permet d'obtenir plusieurs images avec un flou de défocalisation différent. Cette idée n'est pas neuve, cependant aucune réalisation d'imageur chromatique estimant la profondeur sur des scènes réelles n'a été proposé. Dans le principe, ce procédé est proche du MA-DFD. Cependant, une des difficultés est que les images produites par l'imageur chromatique proviennent de scènes qui ne sont pas identiques, contrairement au cas MA-DFD, mais partiellement corrélées. C'est pourquoi, nous proposons dans ce chapitre un algorithme d'estimation de profondeur dédié au traitement d'images chromatiques (SAC-DFD). Cet algorithme, nommé SAC-GLDFD, utilise le formalisme du chapitre 1, mais

nous proposons ici une modélisation différente de la scène pour tenir compte du chromatisme de l'imageur. Grâce à une deuxième collaboration avec l'IOGS, nous avons pu disposer d'un objectif chromatique qui nous a permis de tester l'algorithme SAC-GLDFD sur des images réelles. La combinaison de cet objectif chromatique et de l'algorithme d'estimation de profondeur constitue le premier imageur chromatique à capacité 3D permettant d'estimer la profondeur sur des scènes réelles. Nous évaluons la précision expérimentale d'estimation de profondeur de ce prototype et nous présentons les cartes de profondeur obtenues sur des images réelles. Ces travaux ont fait l'objet d'une publication partielle dans la conférence COSI en juin 2012.

ORGANISATION DU CHAPITRE

Dans un premier temps, nous présentons les articles de l'état de l'art concernant l'utilisation de pupilles codées et d'optiques chromatiques pour des applications d'augmentation de la profondeur de champ ou d'estimation de profondeur. Puis, nous décrivons un protocole expérimental permettant de caractériser la précision d'estimation de profondeur d'un imageur, utilisable pour les deux solutions optiques étudiées dans ce chapitre. Nous présentons ensuite la caractérisation de la précision d'estimation de profondeur du prototype d'imageur à pupille codée. La dernière partie de ce chapitre est dédiée à la mise en œuvre de la solution chromatique. Nous décrivons l'algorithme d'estimation de profondeur SAC-GLDFD, qui est testé tout d'abord sur des images simulées. Puis, le prototype d'imageur chromatique est décrit ainsi que la caractérisation de sa précision d'estimation de profondeur sur l'axe. Enfin, des cartes de profondeur obtenues avec ce prototype sont présentées et comparées à celles obtenues avec un système actif.

2.1 Présentation des deux solutions optiques

2.1.1 Utilisation d'une pupille codée pour favoriser l'estimation de profondeur

2.1.1.1 État de l'art

Le terme de pupille codée désigne l'ajout d'un masque binaire sur le trajet du faisceau lumineux. Ce masque présente un ensemble de trous qui codent la forme de la FEP. Cette solution optique a été proposée dans la littérature pour différentes applications et en particulier pour favoriser l'estimation de profondeur.

L'utilisation d'un masque codé a initialement été proposée pour l'imagerie utilisant des rayons X et gamma dans les domaines de l'astrophysique ou de l'imagerie médicale. En effet, la déviation des rayons pour ces longueurs d'onde est difficile avec les matériaux traditionnels tels que le verre ou le plastique, et il n'est donc pas possible de focaliser les faisceaux. Les premières solutions pour ce type d'imagerie reposent donc sur l'utilisation d'un sténopé. L'ajout d'un masque codé à la place de l'ouverture ponctuelle a été proposé afin d'améliorer le rapport signal à bruit de l'image. Le traitement classique consiste à effectuer une corrélation de l'image acquise par le masque lui-même [Ables, 1968; Dicke, 1968]. En particulier, [Gottesman et Fenimore, 1989] développent le système MURA (Modified Uniformly Redundant Arrays) dans lequel le masque est optimisé pour avoir un large spectre ce qui permet de maximiser le contenu haute fréquence de l'image restaurée.

Depuis quelques années, les pupilles codées sont utilisées également en photographie pour des applications d'estimation de profondeur ou d'extension de profondeur de champ.

La pupille codée a d'abord été proposée dans un contexte multi-images (MA-DFD). Dans [Hiura et Matsuyama, 1998] les auteurs utilisent une ouverture à deux ou quatre trous

rectangulaires et une série d’images acquises avec différentes positions du plan de mise au point. Dans [Zhou et Nayar, 2009a] est décrit un critère d’évaluation de la performance d’estimation de profondeur pour deux pupilles, qui est dérivé de l’expression du maximum a posteriori (MAP). [Farid et Simoncelli, 1998] proposent d’estimer la profondeur à partir de deux acquisitions avec des pupilles ayant une variation continue de transmission. La première pupille a une intensité gaussienne et la seconde est la dérivée d’une gaussienne. Farid montre que le rapport entre la dérivée de l’image acquise avec la première pupille et l’image acquise avec la deuxième pupille est proportionnelle à la profondeur.

Comme discuté dans le chapitre 1 à la section 1.1.2.6, les techniques issues du MA-DFD nécessitent l’acquisition de plusieurs images, donc une scène statique ou l’utilisation de dispositifs particuliers qui augmentent l’encombrement de l’imageur. C’est pourquoi nous nous intéressons dans la suite plutôt au cas d’une seule acquisition, soit le cas SA-DFD.

Dans un contexte de SA-DFD, plusieurs articles utilisent une pupille codée pour améliorer la qualité de la reconstruction d’une image de grande profondeur de champ. Ainsi, [Veeraraghavan *et al.*, 2007] proposent un masque ayant un spectre large bande afin de conserver le maximum d’information sur la scène et améliorer le résultat de déconvolution du flou de défocalisation. Par ailleurs, [Zhou et Nayar, 2009b] développent un critère d’optimisation de la forme d’une pupille codée qui minimise l’erreur de déconvolution. D’un autre côté la pupille codée est également utilisée pour favoriser l’estimation précise de la profondeur dans le cas mono-image. Contrairement aux références précédentes, l’objectif n’est plus d’étendre le spectre du flou de défocalisation vers les hautes fréquences mais de renforcer sa variation avec la profondeur. Dans [Levin *et al.*, 2007b], le principe est de considérer qu’une modification de la profondeur de l’objet correspond à un changement d’échelle de la FEP de défocalisation, dans laquelle on retrouve la forme du masque codé placé au niveau de la pupille. Si la pupille est choisie pour que la FTM associée possède des zéros, alors un changement de profondeur se traduit par un déplacement de la position de ces zéros dans la FTM et ainsi, une position des zéros devient caractéristique d’une profondeur. Si la présence de zéros dans la FTM induit une perte d’information sur la scène, elle permet une estimation plus précise du flou de défocalisation. Dans [Levin *et al.*, 2007b] la pupille est optimisée en maximisant la distance de Kullback entre les FEP à différentes échelles et des comparaisons qualitatives sont fournies entre les images restaurées avec ou sans pupille codée. Levin ne donne cependant pas la précision d’estimation de profondeur obtenue avec ce type de pupille. Par ailleurs, dans [Morrison, 2011] une pupille codée est construite à partir d’un réseau zoné de Fresnel en amplitude. Cette pupille est choisie en particulier car elle possède différents plans de focalisation ce qui permet à la fois d’avoir une FEP variable avec la profondeur, grâce aux plans de focalisation proches, mais de conserver une information haute fréquence de la scène, grâce aux plans de focalisation éloignés. Les pupilles associées aux articles cités précédemment sont présentées à la figure 2.1.

Dans les références précédentes, les pupilles sont testées sur des images réelles mais la précision d’estimation de profondeur n’est pas souvent évaluée par rapport à une vérité terrain. Par contre, dans [Martinello et Favaro, 2011b] plusieurs des pupilles de la figure 2.1 sont comparées en termes d’estimation de profondeur et de restauration d’image de grande profondeur de champ. Ces comparaisons sont conduites avec des FEP calibrées mais sur des images simulées et la pupille proposée par [Levin *et al.*, 2007b] fait partie de celles permettant d’obtenir les meilleurs performances.

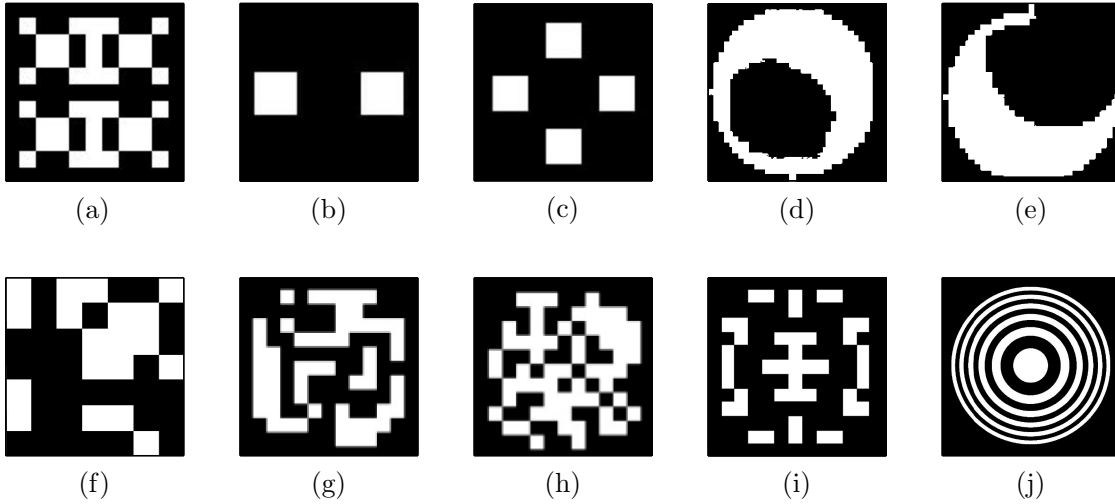


Fig. 2.1: Exemples de pupilles proposées dans la littérature : (a) [Gottesman et Fenimore, 1989], (b) et (c) [Hiura et Matsuyama, 1998], (d) et (e) [Zhou et Nayar, 2009a], (f) [Veeraraghavan *et al.*, 2007], (g) et (h) [Zhou et Nayar, 2009b] pour deux niveaux de bruit différents, (i) [Levin *et al.*, 2007b], (j) [Morrison, 2011].

2.1.1.2 Limitations théoriques

Dans les références traitant de l'estimation de profondeur avec une pupille codée dans le cas SA-DFD, tous les objets visibles de la scène sont placés soit avant soit après le plan de mise au point de l'imageur. En effet, il existe une ambiguïté sur la valeur de la profondeur pour une taille de flou donnée. Cette ambiguïté est illustrée à la figure 2.2 qui représente le module de la taille géométrique de la FEP, définie dans l'équation 1 de l'introduction rappelée ici :

$$\epsilon = 2Rd_{det} \left(\frac{1}{f} - \frac{1}{p} - \frac{1}{d_{det}} \right), \quad (2.1)$$

$$= \frac{f^2 p_0}{F_{/\#}(p_0 - f)} \left(\frac{1}{p_0} - \frac{1}{p} \right) \quad (2.2)$$

où p_0 est la position du plan de mise au point dans l'espace objet, et $F_{/\#}$ le nombre d'ouverture. Cette figure illustre que pour une même taille de flou, il existe deux positions potentielles de l'objet. Ainsi, si un objet est placé avant le plan de mise au point et que le calibrage des FEP a été réalisé de l'autre côté de ce plan, la profondeur estimée de cet objet sera aberrante. La figure 2.3 illustre ce phénomène. Cette image a été prise avec un appareil photographique numérique doté d'une ouverture classique. La famille de FEP potentielles est constituée de gaussiennes d'écart-type variant entre 0.1 et 3. Sur l'image, le plan de mise au point est situé sur l'objet placé au deuxième plan et le deux objets du premier et dernier plan sont flous. Sur la carte de profondeur, l'objet du plan intermédiaire présente bien le flou le plus faible, mais l'objet du premier plan apparaît comme l'objet ayant le flou le plus élevé, donc pourrait être interprété comme l'objet le plus éloigné.

Une solution pour résoudre cette ambiguïté consiste à utiliser une ouverture non symétrique, afin que la FEP ait une forme différente de part et d'autre du plan de mise au point. Ceci permet d'agrandir la région de l'espace où la profondeur peut être estimée. Cependant, dans [Levin *et al.*, 2007b], il est montré qu'une pupille symétrique permet une meilleure discrimination de la profondeur.

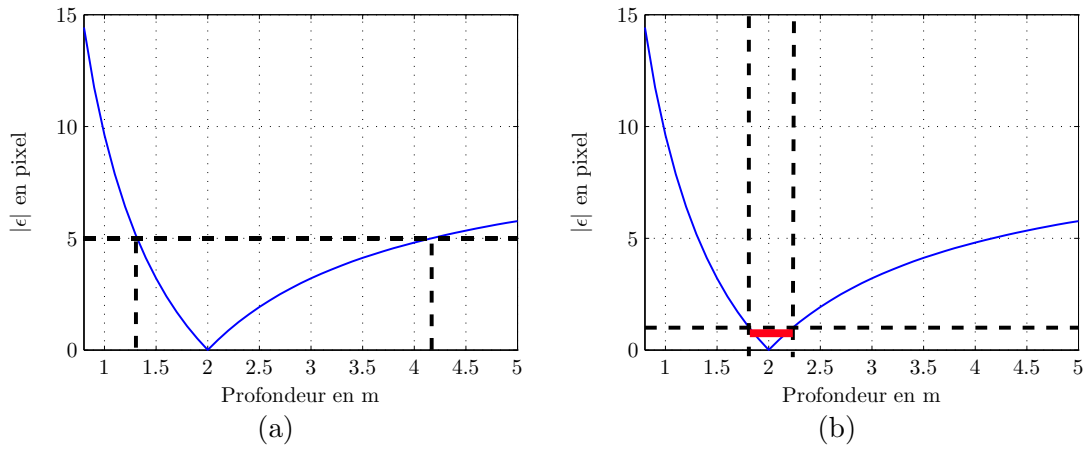


Fig. 2.2: Taille géométrique $|\epsilon|$ de la FEP en pixel pour un imageur de distance focale 20 mm, de nombre d'ouverture 3 et de pixels de $7 \mu\text{m}$, pour une mise au point à 2 m. (a) Illustration de l'ambiguïté sur l'estimation de profondeur. (b) Illustration de la profondeur de la région de l'espace où la profondeur ne peut être estimée précisément. La région en rouge correspond à la zone où la variation de la FEP n'est plus observable.

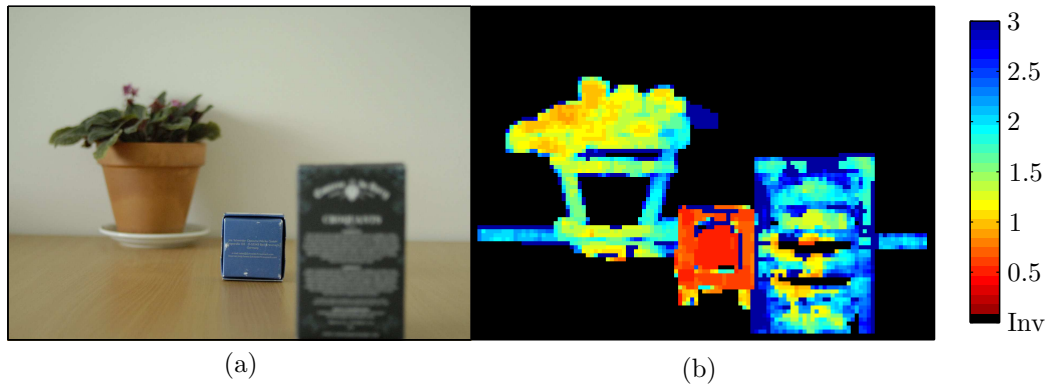


Fig. 2.3: Illustration d'une erreur d'estimation de profondeur liée à l'ambiguïté du flou de défocalisation de part et d'autre du plan de mise au point. L'échelle de couleur correspond à la largeur de l'écart-type des gaussiennes constituant la famille de FEP.

Par ailleurs, il existe une région de l'espace dans laquelle la variation de la FEP n'est pas observable. Cette région correspond à la profondeur de champ, où la dimension de la FEP est inférieure à un pixel. Ainsi, la localisation précise d'un objet dans cette région est impossible comme illustré à la figure 2.2(b). Notons que la courbe de taille de FEP présentée correspond à la taille géométrique et ne tient pas compte de la diffraction. Ainsi, la taille de la FEP dans cette région ne s'annule pas, contrairement à ce qu'indiquent les courbes de la figure 2.2. Néanmoins, comme la résolution de l'imageur est limitée par le détecteur, lorsque la taille de la FEP reste inférieure à la taille du pixel sa variation avec la profondeur n'est pas mesurable.

Il existe donc deux limitations à l'estimation de profondeur mono-image par pupille codée, l'ambiguïté de part et d'autre du plan de mise au point et une région proche du plan

de mise au point, dans laquelle l'estimation est imprécise. Dans la section suivante, nous verrons que le chromatisme permet de résoudre ces limitations.

2.1.2 Utilisation d'une optique chromatique

2.1.2.1 Principe et intérêt

Le chromatisme est une aberration traduisant la variation de la distance focale avec la longueur d'onde, causée par la dispersion de la lumière dans le matériau constituant l'objectif. Comme cette aberration réduit la qualité image, elle est traditionnellement corrigée grâce à l'ajout de lentilles supplémentaires, ce qui augmente l'encombrement, le coût et la difficulté de conception. D'un autre côté, une optique ayant un chromatisme longitudinal non corrigé, combinée à un détecteur couleur, permet d'avoir en une seule acquisition trois images avec différents niveaux de flou. En effet, le détecteur couleur produit plusieurs images dans des bandes spectrales différentes, le plus souvent dans les bandes spectrales des couleurs rouge, verte et bleue. Ces trois bandes sont désignées couramment par les termes de canaux RVB. Puisque la FEP varie spectralement, la taille du flou est différente dans ces trois canaux. Ce phénomène est illustré à la figure 2.4 qui représente la variation de la taille géométrique du flou pour trois canaux RVB d'une optique chromatique. Dans cet exemple les canaux RVB focalisent respectivement à 1.3 m, 2 m et 3 m.

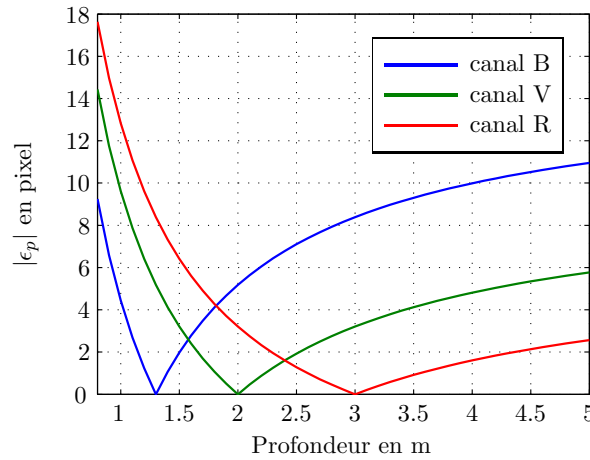


Fig. 2.4: Exemple de variation du diamètre géométrique du flou pour les trois canaux d'une optique chromatique.

La figure 2.4 illustre les avantages d'une optique chromatique par rapport à une optique classique pour l'estimation de profondeur. Tout d'abord, il existe un triplet unique de FEP potentielles pour chaque distance. Ainsi, l'ambiguïté sur la profondeur autour du plan de mise au point disparaît. De plus, dans la configuration de la figure 2.4 lorsque l'un des canaux est net, les deux autres canaux sont flous, donc la variation de la FEP est observable sur ces canaux. Cette caractérisation permet d'éliminer la zone de forte incertitude autour du plan de mise au point que nous avons décrit précédemment.

De plus, contrairement aux systèmes optiques complexes proposés dans la littérature pour produire simultanément plusieurs images avec des flous différents [Hiura et Matsuyama, 1998; Green *et al.*, 2007; Nagahara *et al.*, 2010], l'utilisation du chromatisme permet de simplifier le système optique et de réduire son coût et son encombrement. Enfin, l'acquisition des trois canaux RVB est instantanée, l'estimation de profondeur n'impose donc pas de contrainte sur la scène.

2.1.2.2 État de l'art

L'estimation de profondeur en exploitant le chromatisme longitudinal d'un objectif a été proposée dans [Garcia *et al.*, 2000]. Cependant, la mise en œuvre expérimentale de ce concept est très limitée et concerne uniquement les bords de plages noirs/blancs. Le chromatisme a par contre été exploité dans le domaine de l'extension de la profondeur de champ de l'image, notamment sur des images en noir et blanc [Robinson *et al.*, 2008]. Dans [Milgrom *et al.*, 2010], les auteurs utilisent une optique chromatique pour imager un code barre sur une grande profondeur de champ, en s'arrangeant pour avoir toujours l'un des canaux net. Dans [Cossairt, 2010], l'aberration chromatique est utilisée pour rendre la FEP polychromatique invariable avec la profondeur lorsqu'elle est vue par un capteur noir et blanc. Ainsi, une seule déconvolution est nécessaire pour augmenter la profondeur de champ de l'image. Dans [Lim *et al.*, 2008; Guichard *et al.*, 2009], le chromatisme est utilisé pour augmenter la profondeur de champ d'images couleurs. Le principe est d'avoir toujours un des canaux net et de transférer l'information haute fréquence de ce canal sur les autres canaux flous. Un critère de netteté sert à sélectionner le canal le plus net ce qui permet de classer grossièrement les objets dans les trois classes de profondeur : proche intermédiaire, ou lointaine. Une approche d'estimation de profondeur est mentionnée dans le brevet de l'entreprise DXO [Chanas *et al.*, 2005] en utilisant un calibrage de la courbe de netteté, mais à ce jour aucun dispositif concret ni aucune carte de profondeur expérimentale utilisant cette technique n'a été présentée.

Une autre approche exploitant les canaux RVB du détecteur consiste à découper la pupille avec des filtres de couleurs, ce qui introduit un décalage des images RVB sur le détecteur [Bando *et al.*, 2008; Maik *et al.*, 2010]. Ceci revient à un effet stéréoscopique et la mesure du décalage permet d'estimer la profondeur.

Très récemment, dans la référence [Chakrabarti et Zickler, 2012] est proposée l'utilisation d'une pupille de largeur différente pour le canal vert par rapport aux canaux R et B ajouté à un objectif achromatique. L'image verte a alors une profondeur de champ plus grande que les deux autres canaux. Chakrabarti obtient ainsi trois images de la même scène avec un niveau de flou différent entre les canaux et estime la profondeur en utilisant une modélisation des gradients de l'image RVB qui sont supposés proportionnels. Cependant, puisque ce système n'utilise qu'un seul plan de mise au point, ce dernier possède les mêmes inconvénients qu'un imageur monovoie que nous avons présenté à la section 2.1.1.2, liés à l'ambiguïté de part et d'autre du plan de mise au point et à la profondeur de champ.

2.2 Protocole expérimental

Dans cette section, nous décrivons le protocole expérimental utilisé pour mesurer la précision d'un imageur quelconque combiné à un algorithme d'estimation de profondeur fonctionnant sur des fenêtres à partir d'une famille de FEP calibrées. Ce protocole est utilisé pour les deux prototypes d'imageur favorisant l'estimation de profondeur étudiés dans ce chapitre.

2.2.1 Calibrage des FEP

L'algorithme d'estimation de profondeur que nous proposons permet de sélectionner la FEP la plus vraisemblable dans une famille de FEP potentielles, chacune associée à une profondeur. Le premier enjeu est donc de construire cette famille de FEP. Une première solution peut consister à modéliser ces FEP par exemple en utilisant des modèles simples

basés sur l'optique de Fourier. Si les plans des lentilles constituant l'imageur sont connus, les FEP peuvent aussi être simulées en utilisant un logiciel de conception optique tel que **Zemax**, **Oslo** ou **CodeV**. Cependant, il est rare de disposer des plans des objectifs du commerce et de plus des erreurs de fabrication peuvent entraîner l'apparition d'aberrations qui ne seront pas modélisées dans ces logiciels. La deuxième solution consiste à calibrer les FEP expérimentalement. Nous avons choisi cette solution et nous présentons plusieurs méthodes possibles de calibrage de FEP dans les paragraphes suivants.

2.2.1.1 Acquisition de l'image d'un point source

Une première méthode d'estimation de FEP consiste à acquérir l'image d'un point source placé à différentes distances de l'imageur. Le point source est réalisé en plaçant une plaque métallique trouée devant une source de lumière blanche, qui excite ainsi un large spectre de longueurs d'onde. Le « trou source » ainsi obtenu doit être suffisamment petit pour que son image géométrique soit plus petite que la taille de la FEP. Cependant, cette contrainte intervient surtout pour le calibrage au niveau du plan de mise au point car lorsque la source s'éloigne de ce plan, la taille de l'image géométrique du trou source est négligeable devant la tache de défocalisation.

2.2.1.2 Méthode de Levin [Levin et al., 2007b]

La méthode présentée dans [Levin *et al.*, 2007b] consiste à acquérir une série d'images de la mire présentée à la figure 2.5(a), qui contient un grand nombre de hautes fréquences spatiales. Le principe est de faire une acquisition de la mire au niveau du plan de mise au point, notée \mathbf{y}_{p_0} , qui est assimilée à la scène, puis une acquisition à une profondeur p , notée \mathbf{y}_p . Les deux images sont recalées et la FEP h_p est estimée en minimisant le critère des moindres carrés suivant :

$$\widehat{h_p} = \underset{h_p \geq 0}{\operatorname{argmin}} \|\mathbf{y}_p - h_p * \mathbf{y}_{p_0}\|^2, \quad (2.3)$$

Le recalage vise à compenser le facteur d'échelle entre les deux images, il est réalisé image par image, grâce à des algorithmes déjà développés au DTIM. La résolution de l'équation (2.3) est réalisée à l'aide de la fonction `lsqlin` de **Matlab** en contraignant la FEP à des valeurs positives ou nulles.

2.2.1.3 Méthode de Delbracio [Delbracio et al., 2011]

Cette méthode prévoit également d'estimer la FEP par un critère des moindres carrés, mais une seule acquisition de la cible présentée à la figure 2.5(b) suffit. En effet, au lieu d'utiliser une acquisition dans le plan de mise au point comme référence, Delbracio utilise la scène théorique ayant servi à générer la cible, recalée par rapport à l'acquisition. Ce recalage est effectué grâce à plusieurs points de repères situés sur les bords de la cible. Un filtrage des hautes fréquences permet d'éviter l'aliasing lors du recalage et une correction de l'illumination est appliquée à l'acquisition afin d'avoir 50 % de pixels blancs et 50 % de pixel noirs, conformément au bruit de Bernoulli ayant servi à générer la mire centrale. Enfin la FEP est estimée par l'optimisation suivante :

$$\widehat{h_p} = \underset{h_p \geq 0}{\operatorname{argmin}} \|\mathbf{y}_p - h_p * \mathbf{x}_{\text{theo,rec}}\|^2, \quad (2.4)$$

où \mathbf{y} est l'acquisition et $\mathbf{x}_{\text{theo,rec}}$ la scène théorique après recalage sur l'image. Les codes permettant de faire cette estimation sont disponibles sur le site Internet ¹. L'estimation

¹http://www.ipol.im/pub/algo/admm_non_blind_psf_estimation/

de la FEP est codée en C ce qui la rend très rapide (une estimation de FEP en quelques secondes). Cette méthode prévoit également l'estimation de la FEP subpixellique, ce qui est intéressant lorsque le détecteur présente un filtre de Bayer qui sous-échantillonne les données.

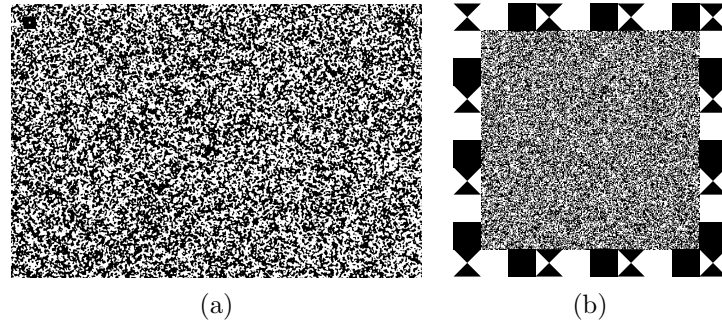


Fig. 2.5: (a) Mire utilisée dans la méthode [Levin *et al.*, 2007b]. (b) Mire utilisée dans [Delbracio *et al.*, 2011].

Notons que [Joshi *et al.*, 2008] propose également une méthode de calibrage de FEP en utilisant une mire adaptée à la modélisation du contour de la scène par une fonction de Heaviside. Cette méthode n'a pas été testée dans ces travaux.

Ces trois méthodes de calibrage ont été utilisées au cours de cette thèse, mais le plus souvent les FEP sont calibrées en utilisant la méthode de Delbracio, car comme nous le discutons dans la section 2.3.2 elle est entre autre la méthode la plus pratique à mettre en oeuvre.

2.2.2 Acquisition de cibles à une profondeur connue

Une fois les FEP calibrées, l'objectif est d'acquérir l'image d'une scène plane et d'estimer sa profondeur à l'aide d'un algorithme d'estimation de profondeur. Un ensemble de cibles de textures et de couleurs variables a été constitué. Plusieurs de ces cibles sont présentées à la figure 2.6. Chaque cible est positionnée parallèlement au plan du détecteur de l'imageur, à une profondeur connue grâce à un imageur actif tel qu'un télémètre. À partir de l'acquisition d'une cible, l'estimation de profondeur est réalisée sur une région au centre de l'image où la FEP est supposée constante. Cette région est divisée en fenêtres, avec ou sans recouvrement, et la profondeur est estimée sur chacune d'entre elles. Un ensemble de mesures est ainsi obtenu, ce qui permet de caractériser la performance d'estimation de profondeur grâce à une étude statistique des résultats, en comparaison avec la profondeur attendue donnée par le télémètre.

2.2.3 Traitement des résultats

L'estimation de profondeur sur plusieurs fenêtres permet d'avoir N résultats de mesures de profondeurs. La performance d'estimation de profondeur est alors caractérisée à l'aide de ces mesures. Les deux outils statistiques classiquement utilisés sont la moyenne et l'écart-type. Cependant, ces derniers sont sensibles à la présence de valeurs aberrantes, ce qui arrive parfois lors de l'estimation de profondeur sur des scènes réelles comme nous avons pu le constater dans le chapitre 1. La médiane est un outil moins sensible aux valeurs aberrantes et la dispersion des valeurs peut être caractérisée par la moyenne de l'écart absolu à la

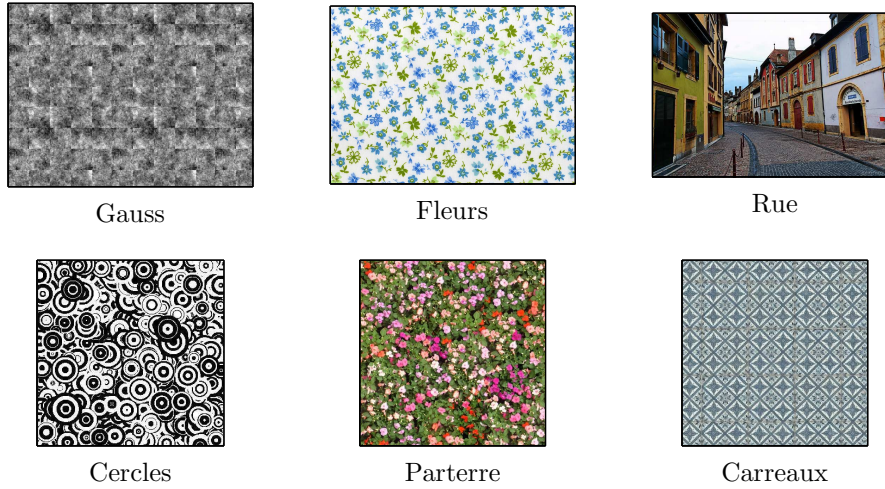


Fig. 2.6: Exemples de cibles utilisées pour évaluer l'estimation de profondeur.

médiane défini par :

$$\sigma_M = \frac{1}{N} \sum_{i=1}^N |\hat{p}_i - M|, \quad (2.5)$$

où M est la valeur médiane. Cette grandeur n'étant pas élevée au carré, elle est moins sensible à la présence de valeurs aberrantes. L'inconvénient de la médiane est qu'elle sélectionne uniquement une des profondeurs de la famille de FEP calibrées, ce qui introduit un biais lorsque la vraie profondeur est située entre deux profondeurs de cette famille.

Enfin, une solution consiste à éliminer les valeurs aberrantes de l'ensemble des résultats. Pour cela, les N résultats sont classés par ordre de valeurs croissantes. En supprimant les valeurs extrémales dans ce classement, il est possible d'éliminer les valeurs aberrantes. Il s'agit alors de décider du pourcentage de valeur supprimées. Dans ce chapitre, cette troisième solution a été utilisée avec un pourcentage total de mesures éliminées de 5%, avec 2.5% des valeurs les plus grandes et 2.5% des valeurs les plus faibles.

2.2.4 Sources d'erreur

L'objectif de ce paragraphe est de considérer les sources d'erreurs d'estimation de profondeur qui ne sont pas liées à l'algorithme lui-même mais au protocole décrit précédemment.

2.2.4.1 Précision de la vérité terrain

Deux systèmes actifs d'estimation de profondeur sont utilisés pour définir la vérité terrain. Un télémètre est utilisé notamment pour le calibrage des FEP et pour la mesure de profondeur sur l'axe. La précision indiquée par le constructeur de cet outil est de l'ordre de 1.5 mm à 30 m avec une portée de 40 m. Le télémètre est fixé le plus près possible de l'imageur, avec la ligne de visée du télémètre parallèle à l'axe optique. Cependant, une erreur d'orientation du télémètre peut se produire. En supposant un écart angulaire de 2° entre l'axe optique et la ligne de visée du télémètre, l'erreur sur la mesure de distance est de moins de 0.1%. Ainsi, l'erreur de mesure en utilisant le télémètre est de l'ordre de 5 mm à 5 m. Cette faible erreur nous incite à utiliser le télémètre comme référence pour l'estimation de profondeur sur l'axe.

Le deuxième système actif utilisé comme référence est la Kinect qui permet d'avoir des cartes de profondeur de la scène imagée. La portée de cet instrument est de 5 m environ avec une précision de l'ordre de 1 mm à 1 m et de 5 cm à 5 m [Chow *et al.*, 2012]. Cet outil, moins précis que le télémètre, est utilisé car il permet d'obtenir une carte de profondeur de référence.

2.2.4.2 Variation de la FEP dans le champ

L'estimation de profondeur est effectuée sur une région de l'image où la FEP est supposée constante. Or la FEP varie continûment dans le champ ce qui introduit des erreurs sur l'estimation de profondeur. Cette erreur est délicate à quantifier théoriquement car la variation de la FEP dépend de l'imageur et de la profondeur. Toutefois, un moyen de limiter cette erreur est d'estimer la FEP pour différentes valeurs du champ et d'observer à partir de quel angle la variation de la FEP n'est plus négligeable.

2.2.4.3 Angle entre le plan du détecteur et le plan de la cible

Les acquisitions sont réalisées à différentes profondeurs par un déplacement soit de la scène soit de l'imageur pour chaque acquisition. Ce déplacement est réalisé manuellement ce qui peut introduire un angle entre le plan de la cible et celui du détecteur. Comme l'estimation de profondeur est moyennée sur une région de l'image où la profondeur et la FEP sont supposées constantes ceci peut créer une augmentation artificielle de l'écart-type des valeurs de profondeur estimées car la cible présente alors un tilt et la profondeur n'est plus constante spatialement. Afin de limiter cette erreur, une ligne de référence est positionnée sur le sol, perpendiculairement au plan de la cible (resp. du détecteur) en suivant un marquage initial présent sur le sol du laboratoire où sont réalisées les expérimentations. L'imageur (resp. la cible) est alors reculé(e) suivant cette ligne. Avec ce procédé, l'erreur sur l'angle entre le plan de la cible et le plan du détecteur est estimée inférieure à $\pm 1^\circ$ à 4 m et de $\pm 2^\circ$ à 1 m. C'est donc pour des profondeurs faibles que cette erreur aura le plus d'impact. De plus, l'effet du tilt dépend de la taille de la région où est estimée la FEP. Par exemple pour avoir une erreur inférieure à 5 mm, la région de l'image où est estimée la profondeur doit correspondre à une région dans l'espace objet inférieure à 30 cm à 1 m et à 60 cm à 4 m.

2.2.4.4 Calibrage de la FEP

L'estimation de profondeur repose sur la sélection d'une FEP parmi une famille de FEP potentielles. Cependant, ces FEP sont obtenues par calibrage et ne sont donc que des estimations des vraies FEP. Une simulation est réalisée afin de quantifier l'effet de l'erreur de calibrage des FEP sur la performance d'estimation de profondeur. Les FEP utilisées ici sont les mêmes que celles utilisées dans le chapitre 1 au paragraphe 1.3.1.1, simulées pour un imageur de distance focale 25 mm, d'ouverture 2.5 et dont les pixels font $5 \mu\text{m}$. Le plan de mise au point est situé à 2 m. Pour plusieurs FEP correspondant à des profondeurs comprises entre 2 et 3 m, une image simulée est générée en concaténant des fenêtres de tailles 21×21 pixels, simulées à partir de scènes vérifiant l'a priori gaussien et avec l'ajout d'un bruit blanc gaussien tel que $\alpha = 0.001$. La figure 2.7 présente les résultats d'estimation de profondeur obtenus par le critère GL décrit dans le chapitre 1 en utilisant deux familles différentes de FEP potentielles. Dans la figure 2.7 (a) la famille de FEP potentielles est constituée des FEP ayant servi à générer les images et dans la figure 2.7 (b), la famille de FEP potentielles est constituée des mêmes FEP mais bruitées avec un bruit blanc gaussien

d'écart-type 0.03 fois la hauteur maximale de la FEP, afin de simuler une erreur de calibrage.

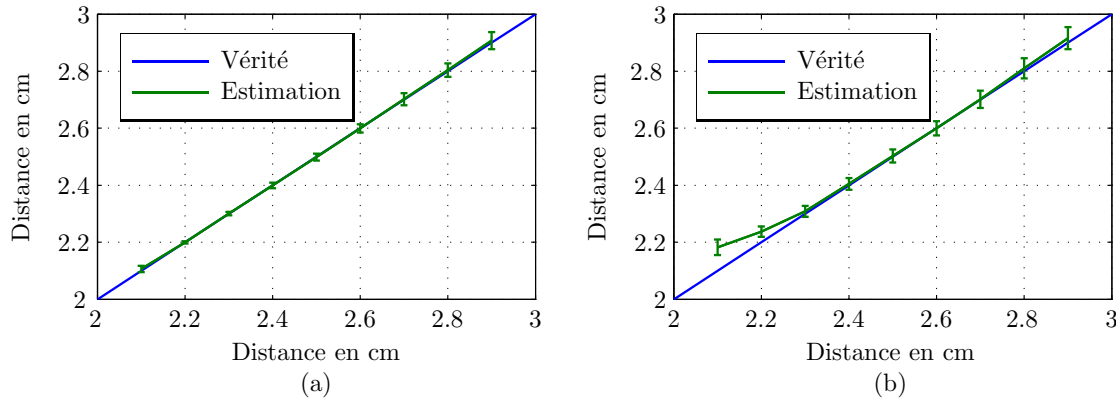


Fig. 2.7: Estimation de profondeur sur des images simulées avec une famille de FEP potentielles constituées respectivement : (a) des FEP théoriques, (b) des FEP théoriques bruitées afin de simuler l'erreur de calibrage.

La figure 2.7 montre qu'une erreur sur les FEP augmente globalement l'écart-type de profondeur et ajoute un biais sur les résultats notamment pour les profondeurs proches du plan de mise au point. Pour cet imageur, cette erreur introduit une augmentation de l'écart-type d'un facteur 2, ce dernier s'élève alors à environ 3 cm.

Outre l'erreur de calibrage de la FEP, les résultats d'estimation de profondeur dépendent du pas de calibrage, comme l'illustre la figure 2.8. Cette figure représente l'histogramme des résultats d'estimation de profondeur de l'algorithme pour trois pas de calibrage Δ : 20 cm, 5 cm et 1 cm, sachant que la première FEP calibrée est celle située à 2 m. La figure (a) correspond à un objet placé à 2.7 m et les figures (b) et (c) à un objet placé à 2.2 m. La famille de FEP potentielles est celle utilisée dans la simulation précédente. Les figures 2.8 (a) et (b) sont obtenues pour des scènes vérifiant l'a priori gaussien et (c) pour une scène naturelle. Le tableau 2.1 présente les valeurs moyennes et les écarts-types obtenus dans chaque cas, en enlevant 5 % des valeurs extrémales.

Moyenne tronquée en m				Écart-type tronqué en cm			
Δ (cm)	20	5	1	Δ (cm)	20	5	1
(a)	2.73	2.70	2.70	(a)	9.5	2.7	2.1
(b)	2.20	2.2	2.20	(b)	0	0	0.65
(c)	2.25	2.25	2.25	(c)	11	8.5	8.5

Tableau 2.1: Moyenne et écart-type tronqués associés aux histogrammes de la figure 2.8.

Dans la figure 2.8(a), la vraie distance est 2.7 m. Pour un pas de calibrage de 20 cm, la vraie profondeur ne fait donc pas partie de la famille de FEP potentielles. Ainsi, les résultats de l'algorithme se répartissent autour de 2.7 m, la moyenne est biaisée de 3 cm et l'écart-type est de l'ordre de 10 cm ce qui correspond à la moitié du pas de calibrage. Pour un pas de 5 cm, la vraie profondeur fait partie des FEP potentielles, le biais et surtout l'écart-type diminuent, cependant les résultats ne se répartissent que sur trois profondeurs. Enfin, pour un pas de 1 cm, une vraie distribution des résultats se forme autour de la vraie

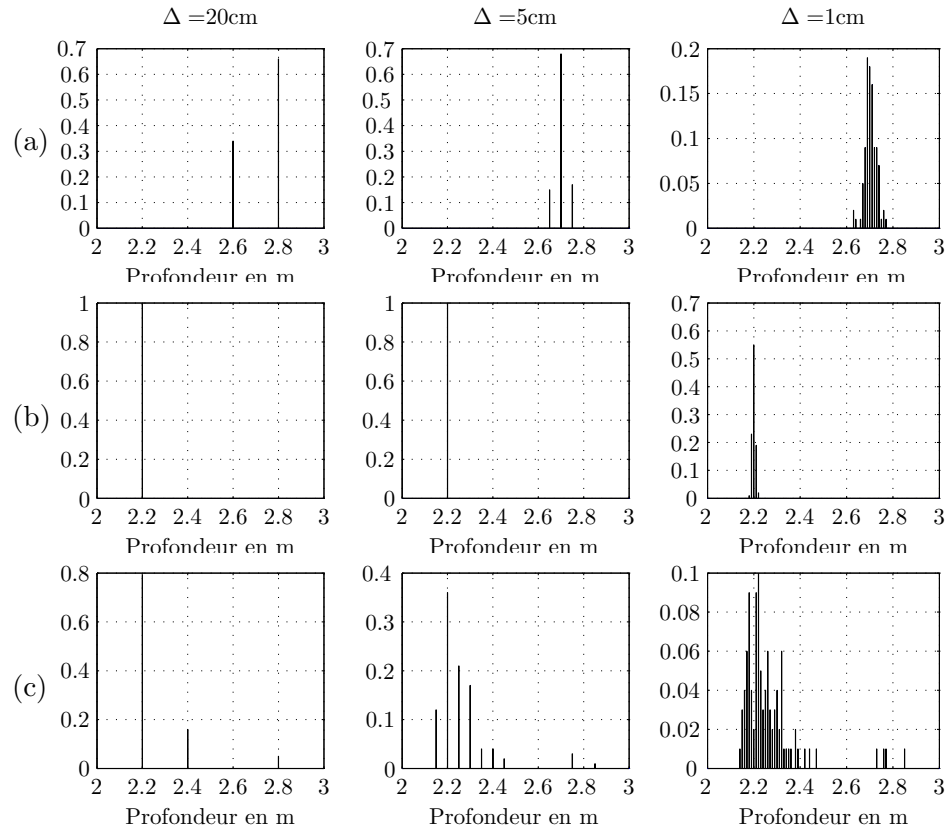


Fig. 2.8: Histogrammes des résultats d'estimation de profondeur obtenus pour trois pas de calibrage. (a) Les scènes sont générées suivant l'a priori gaussien et la profondeur vraie est à 2.7 m. (b) Les scènes sont générées suivant l'a priori gaussien et la profondeur vraie est à 2.2 m. (c) La scène est naturelle et la profondeur vraie est à 2.2 m.

profondeur. Notons que l'écart-type est toutefois proche de la valeur obtenue pour un pas de 5 cm. La figure 2.8(b) est obtenue comme pour la figure (a) mais pour un objet situé à 2.2 m. Les observations qui peuvent être faites sont similaires, si ce n'est que l'écart-type est nul pour le pas de calibrage de 20 cm, du fait que la vraie FEP appartient cette fois-ci à la famille de FEP potentielles. De plus, avec un pas de 1 cm la dispersion des résultats est moins importante.

Dans la figure 2.8(c), il ne s'agit plus d'une scène respectant l'a priori gaussien mais d'une scène naturelle. Dans ce cas l'écart-type des résultats est plus important et des valeurs aberrantes apparaissent. Ainsi, en conservant tous les résultats, l'écart-type pour un pas de 1 cm est de 12.5 cm alors que l'écart-type sans valeurs aberrantes est de 8.5 cm. Ceci illustre que les valeurs aberrantes influencent la valeur de l'écart-type. Dans ce chapitre, nous choisissons de tronquer un certain pourcentage des valeurs extrémales, pour traiter le cas d'images issues de scènes naturelles, afin que les résultats statistiques ne soient pas biaisés par des valeurs aberrantes. On se souviendra cependant que des valeurs aberrantes apparaissent lors de l'estimation de profondeur sur des images naturelles.

En conclusion, le calibrage de la FEP influence les résultats d'estimation de profondeur. Comme on pouvait s'y attendre, plus le pas de calibrage est fin, plus la distribution des résultats se fait sur un grand nombre de profondeurs, ce qui permet d'estimer correctement la statistique des résultats. En contrepartie, un pas trop petit implique un grand nombre

de FEP à calibrer et à tester dans l'algorithme, ce qui augmente le temps de calibrage et de traitement. Enfin, le pas nécessaire à l'obtention d'une bonne répartition des résultats dépend de la profondeur. Une solution serait d'adapter le pas de calibrage en fonction de celle-ci, mais il faudrait alors utiliser un a priori sur la précision d'estimation en fonction de la profondeur. Or, l'objectif de ce chapitre est justement d'étudier expérimentalement cette dépendance, aucun a priori n'est volontairement introduit et le pas de calibrage est le même pour toutes les profondeurs. L'estimation de profondeur sur une scène naturelle présente un écart-type plus important que lorsque la scène vérifie l'a priori gaussien. Dans cette simulation il est de l'ordre de 10 cm.

2.2.4.5 Bilan

La revue des sources d'erreurs expérimentales a montré que la principale source d'erreur provient du calibrage des FEP. De plus, le pas de calibrage a une influence sur les statistiques des résultats et doit être suffisamment fin pour que l'étude statistique soit fiable. L'étude d'un exemple d'estimation de profondeur sur une scène naturelle a donné un écart-type de l'ordre de 10 cm, ce qui dépasse largement les sources d'erreur liées au positionnement de la cible par rapport à l'imageur. Étant donné cet écart-type, un pas de calibrage de l'ordre de quelques cm semble être une bonne valeur. C'est pourquoi dans ce chapitre, les FEP sont calibrées avec un pas fixe de 5 cm. Nous définissons alors pour chaque imageur une région de l'espace dans laquelle la variation de la FEP peut être négligée, y compris celle liée à une erreur de tilt.

2.3 Étude d'un imageur à pupille codée

Dans cette section nous présentons le prototype d'imageur à pupille codée qui a été réalisé au cours de cette thèse ainsi que plusieurs résultats de FEP calibrées. Puis les performances d'estimation de profondeur de ce prototype sont étudiées sur l'axe pour différentes cibles. Enfin, nous présentons des résultats d'estimation de profondeur sur des images complètes, en comparaison avec les cartes de profondeur produites par un système actif, à savoir la Kinect.

2.3.1 Réalisation d'un prototype

Avec l'aide de T. Avignon de l'IOGS, une pupille codée a été insérée à l'intérieur d'un appareil photographique numérique Nikon D200 muni d'un objectif de 35 mm. La pupille a la même forme que celle proposée par [Levin *et al.*, 2007b] présentée à la figure 2.1(i) car cette pupille a montré les meilleures performances d'estimation de profondeur sur des images simulées [Martinello et Favaro, 2011b]. La figure 2.9 présente l'appareil photographique Nikon muni de la pupille codée.

2.3.2 Calibrage des FEP

La première étape consiste à calibrer les FEP. Nous avons présenté dans la section 2.2.1 trois méthodes de calibrage potentielles. Nous présentons ici les résultats de FEP calibrées suivant ces trois méthodes. La mise au point de l'imageur à pupille codée est placée à 1.9 m et la figure 2.10 présente des exemples de FEP obtenues sur l'axe avec les trois méthodes de calibrage. La figure 2.11 présente le module des transformées de Fourier des trois FEP à la profondeur 3.5 m.



Fig. 2.9: Imageur à pupille codée.

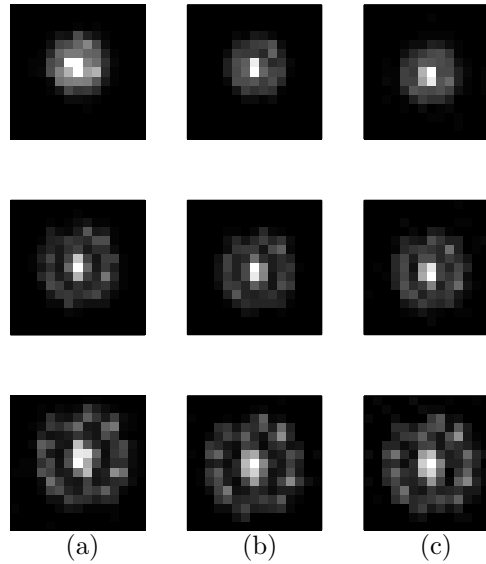


Fig. 2.10: Calibrage des FEP par : (a) un point source, (b) la méthode de [Levin *et al.*, 2007b] et (c) la méthode de [Delbracio *et al.*, 2011]. De haut en bas les profondeurs sont 2.7, 3.2 et 3.5 m.

L'allure des FEP obtenues avec les trois méthodes de calibrage sont très similaires. De plus, les modules des transformées de Fourier des pupilles calibrées avec les trois méthodes de calibrage, autrement dit les Fonction de Transfert de Modulation (FTM) en intensité se superposent.

La méthode du point source est plus rapide du point de vue traitement, car l'acquisition donne directement la FEP, mais du point de vue expérimental, elle demande de faire des acquisitions dans une pièce sombre et de modifier le temps d'acquisition ou la sensibilité de l'appareil au cours des acquisitions, car l'intensité de la FEP décroît lorsqu'elle s'étend avec la profondeur. Dans cette méthode le réglage de l'appareil est finalement différent entre le calibrage et les acquisitions, ce qui risque de modifier la FEP puisque le gain des pixels est différent.

La méthode de Levin demande de faire une acquisition au niveau du plan de mise au point et une erreur sur cette acquisition induit une erreur sur toutes les FEP. De plus, l'absence de points de repère sur la mire rend l'étape de recalage plus délicate et longue.

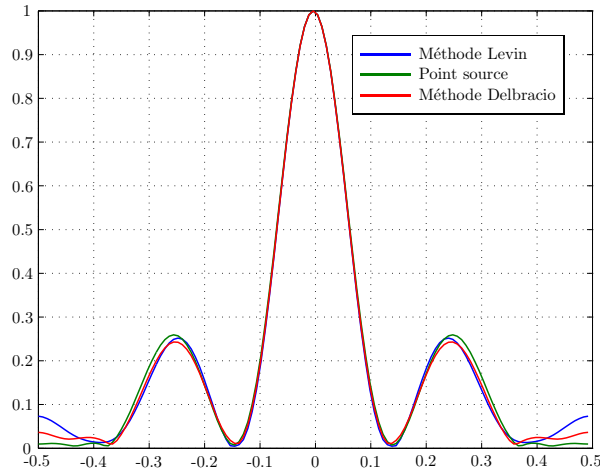


Fig. 2.11: Comparaison des FTM en intensité des pupilles calibrées suivant les différentes méthodes de calibrage pour une profondeur de 3.5 m.

Enfin, pour des flous très importants l'image de la cible devient uniformément grise, ce qui rend l'estimation de la FEP impossible.

La méthode de Delbracio ne demande qu'une seule acquisition, le recalage se fait automatiquement grâce aux points de repères et le traitement d'estimation de FEP est rapide. Dans le cas de l'estimation de FEP défocalisées, la détection de ces points est plus délicate car l'image de la cible est floue. La méthode de Delbracio est prévue pour l'estimation d'une FEP pour des distances proches du plan de mise au point, l'image de la cible étant alors assez nette pour que l'estimation automatique de la position des points servant de repères soit possible. Nous avons donc modifié le code associé à la référence [Delbracio *et al.*, 2011] pour prévoir un positionnement manuel des points de repère de la cible. De plus, cette méthode permet d'estimer la FEP à l'échelle subpixelique ce qui est un avantage supplémentaire de cette méthode, car elle peut nous permettre de calibrer la FEP à l'échelle du détecteur. En contrepartie, pour des variations de profondeur très importantes, il est nécessaire d'utiliser des cibles imprimées à des échelles différentes car les auteurs indiquent que l'image de la cible doit être de l'ordre de la centaine de pixel.

La méthode de Delbracio est donc la méthode la plus pratique expérimentalement et rapide du point de vue traitement. De plus, elle offre un potentiel de calibrage subpixelique, ce qui permet d'estimer la FEP sous-échantillonnée à cause du filtre Bayer, c'est pourquoi cette technique est utilisée dans la suite de ce chapitre.

L'estimation de profondeur ayant lieu sur une fenêtre centrée de l'image, la dimension de cette fenêtre doit correspondre à une région où la FEP est supposée invariante. Afin de définir la taille de cette région, les FEP sont calibrées hors axe en positionnant la mire dans différentes positions de l'image qui sont caractérisées par des positions angulaires (θ_h, θ_v) par rapport à l'axe optique dans les directions horizontale et verticale et pour trois profondeurs. La figure 2.12 présente plusieurs FEP obtenues. Les FEP varient très peu dans le champ. Ainsi, l'objectif à pupille codée, construit à partir d'un objectif du commerce de bonne qualité, présente très peu d'aberrations de champ. Nous faisons donc l'hypothèse que la FEP calibrée au centre peut être utilisée sur toute l'image.

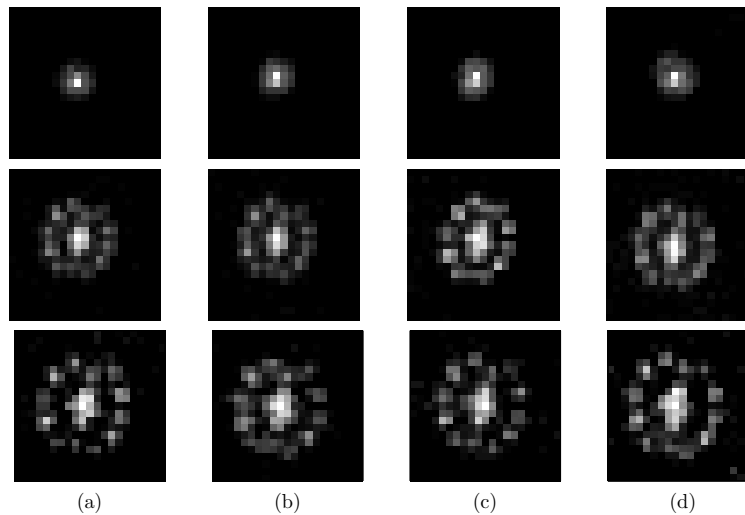


Fig. 2.12: Variation des FEP dans le champ pour trois profondeurs (de haut en bas 2 m, 3 m et 4 m). (a) sur l'axe, (b) $(\theta_h, \theta_v) = (0^\circ, 2^\circ)$, (c) $(\theta_h, \theta_v) = (0^\circ, 10^\circ)$, (d) $(\theta_h, \theta_v) = (10^\circ, 7^\circ)$. Le coin en haut à droite de l'image correspond aux positions angulaires $(\theta_h, \theta_v) = (20^\circ, 13^\circ)$.

2.3.3 Mesures de distance sur l'axe

Le calibrage des FEP est effectué pour une mise au point à 1.6 m entre 1.6 m et 4 m avec un pas de 5 cm en utilisant la méthode de Delbracio. Nous utilisons les images brutes produites par l'imageur au format .raw. L'image brute est séparée en quatre canaux R, V₁, V₂ et B correspondant aux quatre filtres du schéma Bayer. L'algorithme GLDFD est utilisé sur un des canaux verts de l'image qui correspond donc à une version sous-échantillonnée d'un facteur 2 de l'image produite sur le détecteur. Dans un premier temps, la famille de FEP potentielles sont celles calibrées sur ce même canal, également sous-échantillonnées d'un facteur 2.

Variation de la cible

L'estimation de profondeur est ensuite réalisée sur différentes cibles. La figure 2.13 présente les performances d'estimation de profondeur pour quatre des cibles de la figure 2.6 imprimées sur une feuille de format A4. Ces cibles sont les cibles désignées par les noms « Gauss », « Cercles », « Parterre » et « Carreaux ». L'estimation de profondeur est réalisée à l'intérieur d'une région centrée de taille 140×200 pixels sur l'un des deux canaux verts. Cette région correspond à une position angulaire maximale de $(\theta_h, \theta_v) = (3^\circ, 4^\circ)$. L'erreur de parallélisme entre la cible et l'imageur est de l'ordre de quelques mm, ce qui est négligeable devant le pas de calibrage de 5 cm. Les fenêtres de l'algorithme sont de taille 25×25 pixels avec un recouvrement de 50%. 150 résultats d'estimation de profondeur sont ainsi obtenus et l'écart-type et la moyenne sont calculés à partir de ces résultats en enlevant 5% des valeurs extrémales.

Pour les quatre cibles, l'estimation de profondeur présente un écart-type faible, variant entre 5 cm et 10 cm, et un biais négligeable pour des profondeurs comprises entre 1.8 m et 2.8 m. Pour de faibles profondeurs un biais apparaît qui peut s'expliquer par l'effet de la profondeur de champ. En effet, les FEP calibrées dépassent 2 pixels uniquement à partir de 1.8 m, donc entre 1.6 m et 1.8 m aucune variation de la FEP n'est observée ce qui explique que les erreurs sur l'estimation de profondeur soient plus importantes dans

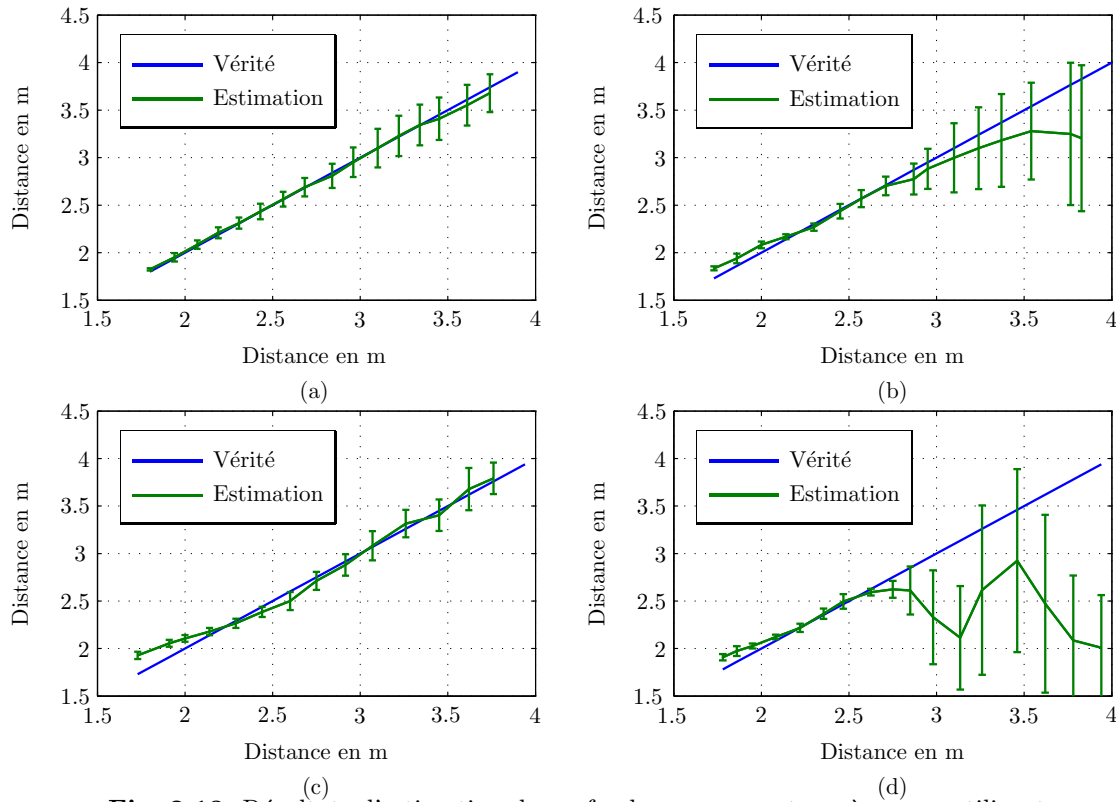


Fig. 2.13: Résultats d'estimation de profondeur pour quatre scènes, en utilisant l'imageur à pupille codée. (a) Cible « Gauss », (b) Cible « Cercles », (c) Cible « Parterre », (d) Cible « Carreaux ».

cette région. Des différences de comportement sont observées suivant les cibles pour des profondeurs dépassant 2.8 m. Pour les cibles « Gauss » et « Parterre » (figure (a) et (c)), le biais reste négligeable jusqu'à 3.7 m. Par contre, après environ 2.8 m l'estimation de profondeur avec la cible « Cercles » présente un écart-type et un biais plus important (figure (b)) et l'estimation de la profondeur avec la cible « Carreaux » dérive complètement (figure (d)). Ceci peut s'expliquer par les différences de textures des cibles. La cible « Gauss » est une concaténation de fenêtres générées suivant l'a priori gaussien, il n'est donc pas surprenant qu'elle donne de bons résultats d'estimation de profondeur. La cible « Parterre » présente une texture relativement isotrope et l'algorithme GLDFD fonctionne correctement. Les scènes « Cercles » et « Carreaux » respectent moins l'a priori gaussien isotrope ce qui peut expliquer une performance moins bonne d'estimation de profondeur.

Comparaison des performances entre pupille codée et pupille classique

Dans ce paragraphe, les performances d'estimation de profondeur obtenues avec un objectif d'ouverture classique sont comparées à celles obtenues avec l'objectif à pupille codée. L'objectif classique a les mêmes caractéristiques optiques que l'objectif à pupille codée, si ce n'est que son ouverture est quasi-circulaire². Son nombre d'ouverture est réglé sur 3.2 afin d'avoir le même diamètre que la pupille codée. Les deux objectifs sont successivement utilisés dans l'appareil photo Nikon avec une mise au point située à 1.6 m. Le temps d'intégration est adapté pour avoir le même rapport signal à bruit dans les images acquises avec les deux

²En réalité l'ouverture a une forme hexagonale.

objectifs. La figure 2.14 présente les résultats d'estimation de profondeur obtenus pour la cible « Gauss » avec les deux objectifs.

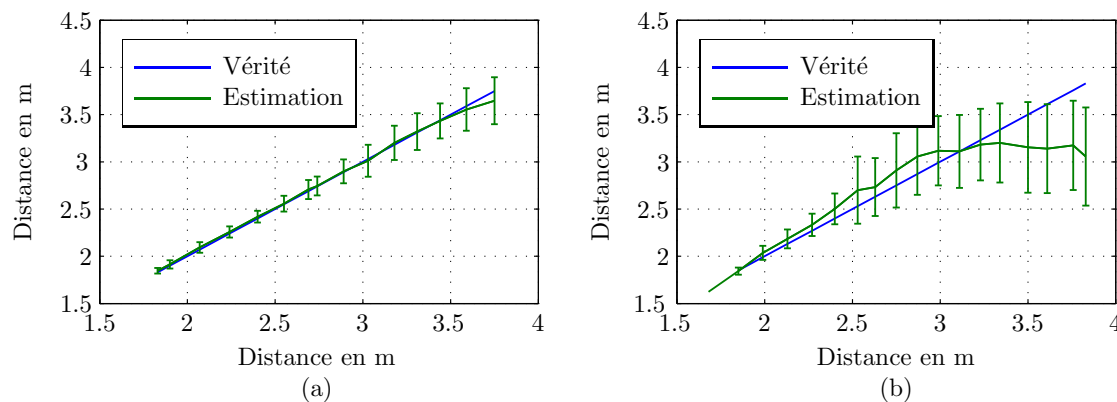


Fig. 2.14: (a) Performances d'estimation de profondeur pour l'imageur : (a) muni d'une pupille codée. (b) muni d'une pupille classique.

L'imageur à pupille codée permet d'estimer la profondeur pour des distances comprises entre 1.6 et 3.7 m, alors que pour la pupille classique l'écart-type augmente fortement à partir de 2.5 m et les valeurs moyennes dérivent complètement à partir de 3 m. Dans la région où l'estimation fonctionne pour les deux imageurs, l'écart-type de l'imageur à pupille classique est également plus important. Ainsi, l'imageur à pupille codée permet d'avoir une meilleure précision et d'estimer la profondeur sur une plus grande distance que l'imageur classique.

Prise en compte de l'effet du schéma Bayer

Dans les tests précédents, l'algorithme GLDFD est utilisé en négligeant l'effet du schéma Bayer. En effet, seul un des deux canaux verts est utilisé pour le calibrage des FEP et l'estimation de profondeur. L'algorithme est appliqué sur une fenêtre de taille 25×25 pixels de ce canal, ce qui correspond finalement à n'exploiter qu'un pixel sur quatre d'une fenêtre de taille 50×50 pixels à l'échelle de l'image brute produite par le détecteur. Or il est possible de modéliser le sous-échantillonnage lié au schéma Bayer, comme discuté dans le chapitre 1 section 1.5.3, et d'utiliser tous les pixels verts de la fenêtre, soit finalement un pixel sur deux. Ceci nécessite alors de connaître la FEP à l'échelle du détecteur. Une nouvelle famille de FEP potentielles est donc construite, en utilisant la méthode de Delbracio subpixelique pour les profondeurs proches du plan de mise au point, ou en interpolant d'un facteur 2 les FEP sous-échantillonnées lorsque celles-ci sont suffisamment larges. La figure 2.15 présente les résultats d'estimation de profondeur obtenus en utilisant les FEP estimée à l'échelle du détecteur vis-à-vis de ceux obtenus précédemment pour la cible « Gauss ».

La figure 2.15 montre que la performance d'estimation de profondeur en modélisant le schéma Bayer et en utilisant tous les pixels verts du détecteur améliore les performances d'estimation de profondeur mais cette amélioration reste légère. L'écart-type diminue et le biais apparaissant aux grandes profondeurs disparaît. La prise en compte du schéma Bayer permet d'exploiter tous les pixels verts du détecteur et non plus un sur deux, ce qui augmente le nombre de données utilisées et améliore théoriquement la précision d'estimation de profondeur. D'un autre côté, le calibrage de la FEP à l'échelle du détecteur ou par interpolation introduit potentiellement plus d'erreurs sur la famille de FEP ce qui peut expliquer que le gain observé sur les performances soit limité.

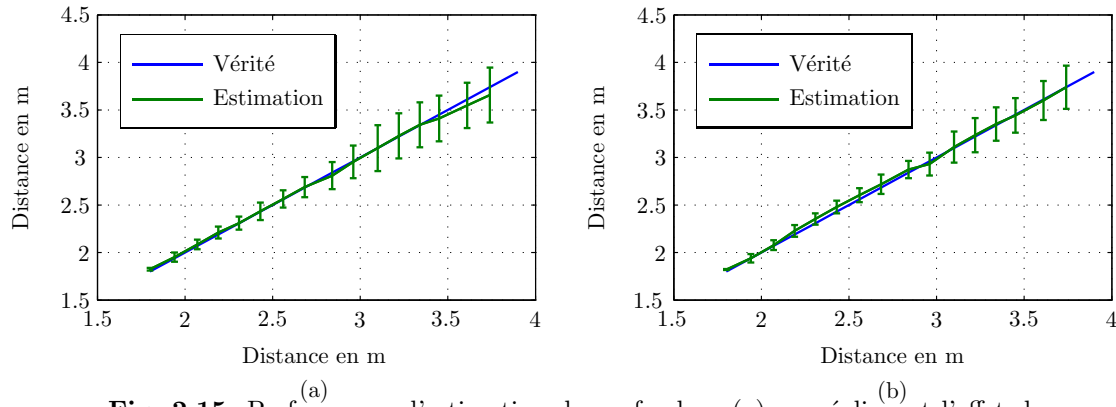


Fig. 2.15: Performance d'estimation de profondeur (a) en négligeant l'effet du schéma Bayer. (b) En prenant en compte l'effet du schéma Bayer.

2.3.4 Cartes de profondeur

Dans les articles de la littérature traitant de la pupille codée pour favoriser l'estimation de profondeur les cartes de profondeur estimées sont présentées sans vérité terrain. Dans le paragraphe suivant les cartes de profondeurs obtenues par l'algorithme GLDFD sont comparées à celles fournies par un système actif, à savoir la Kinect diffusée par Microsoft.

2.3.4.1 Comparaison de cartes de profondeur avec un système actif

La Kinect est positionnée à proximité de l'appareil photographique Nikon à pupille codée et les acquisitions sont réalisées de manière quasi-simultanée sur les deux appareils. La figure 2.16 présente un exemple d'image issue du Nikon et deux cartes de profondeur obtenues respectivement avec la Kinect et avec l'algorithme GLDFD.

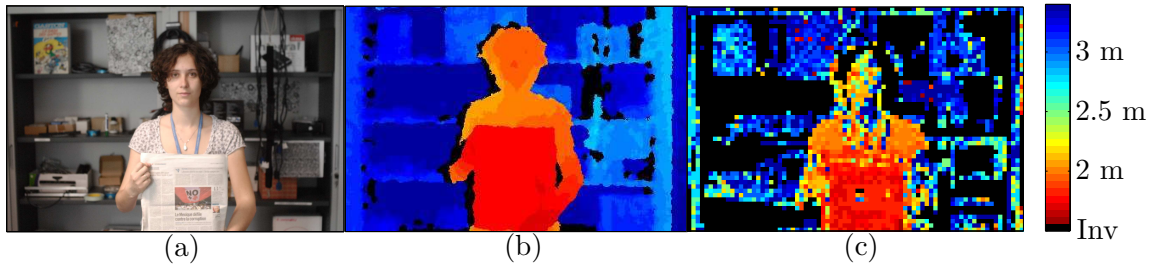


Fig. 2.16: (a) Images acquises avec l'appareil photographique Nikon. Carte de profondeur obtenue : (b) avec la Kinect et (c) en utilisant l'algorithme GLDFD. Les régions en noir correspondent aux régions homogènes rejetées en pré-traitement.

La carte de profondeur obtenue par l'algorithme GLDFD est cohérente avec celle de la Kinect, même si elle est plus bruitée et plus lacunaire, à cause des zones homogènes. Ces résultats illustrent la capacité de l'imageur à ouverture codée à fournir une image de profondeur dans un environnement intérieur.

2.3.4.2 Scènes extérieures

Un défaut de la Kinect est qu'elle ne peut pas être utilisée à l'extérieur, du fait du rayonnement infrarouge du Soleil qui perturbe le signal infrarouge émis par la Kinect. Au contraire, le prototype à pupille codée peut estimer la profondeur, comme le montrent

les figures 2.17 et 2.18 qui présentent les cartes de profondeurs obtenues pour quelques acquisitions. La première acquisition de la figure 2.17 correspond à un mur dont le plan

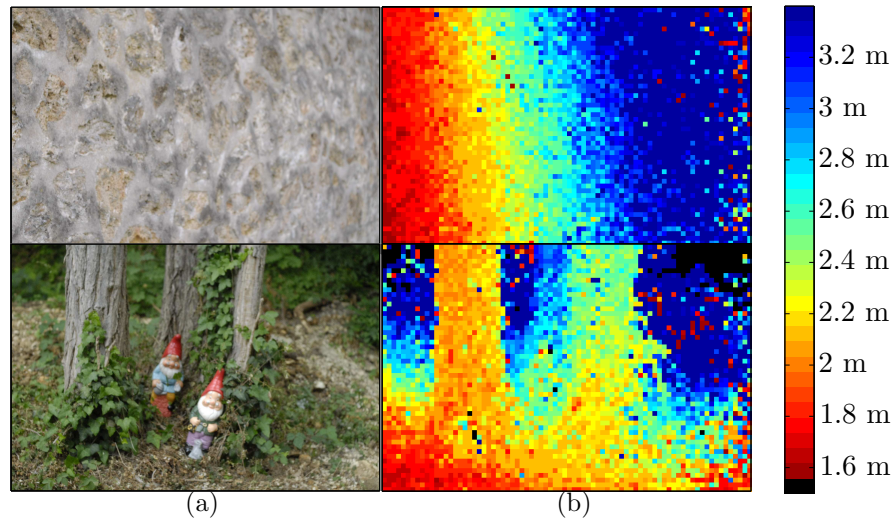


Fig. 2.17: (a) Acquisitions en extérieur avec l'appareil à pupille codée. (b) Cartes de profondeurs obtenues avec l'algorithme GLDFD.

est incliné par rapport au plan du détecteur. La variation continue de profondeur apparaît nettement dans la carte de profondeur obtenue avec l'algorithme GLDFD. La deuxième acquisition présente une scène avec différents niveaux de profondeur qui apparaissent dans la carte de profondeur. Sur l'acquisition de la figure 2.18, réalisée par G. Le Besnerais, un

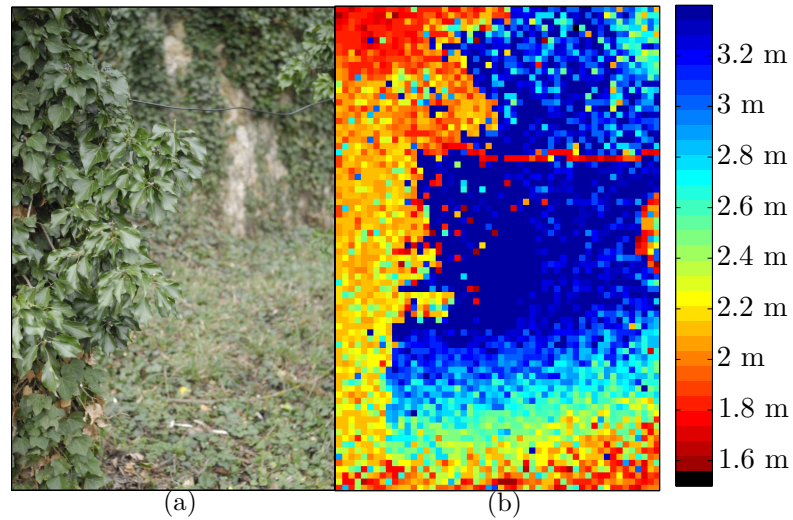


Fig. 2.18: (a) Acquisition en extérieur avec l'imageur à pupille codée. Un câble est tendu entre les deux arbres. (b) Carte de profondeur obtenue l'algorithme GLDFD. Le câble apparaît nettement dans la carte de profondeur.

câble est tendu entre deux arbres. Ce câble, à peine visible dans l'image, apparaît dans la carte de profondeur, ce qui montre la sensibilité de l'algorithme GLDFD pour ce type

d'objet, ce qui est particulièrement intéressant pour des applications de détection d'obstacles pour des appareils autonomes tels que des drones.

2.3.5 Bilan

Dans cette section, nous avons pu mesurer la précision d'estimation de profondeur du prototype à pupille codée. Sur les quatre cibles testées, l'estimation de profondeur a un biais négligeable et un écart-type variant entre 5 et 10 cm pour des distances comprises entre 1.8 et 2.8 m et avec un calibrage des FEP tous les 5 cm. Certaines cibles permettent même d'estimer la profondeur jusqu'à 4 m. Pour ces cibles la précision expérimentale d'estimation de profondeur est de l'ordre de 5 cm à 2 m et de 20 cm à 3.5 m.

La réalisation d'un prototype d'imageur à pupille codée a permis tout d'abord de valider expérimentalement l'algorithme GLDFD présenté au chapitre 1 sur des images réelles avec une vérité terrain. La variété des cibles testées et des conditions d'acquisition utilisées a permis de faire une première évaluation de la robustesse de l'algorithme GLDFD, en particulier vis-à-vis des textures naturelles.

De plus, les cartes de profondeur obtenues sont cohérentes avec un capteur actif en intérieur et les essais de ce type d'imageur que nous avons menés en extérieur ont permis d'obtenir des cartes de profondeurs fidèles à la scène et en particulier une bonne sensibilité pour des objets de taille réduite tels que des câbles. Ces résultats semblent intéressants pour des applications à la navigation autonome d'appareils tels que des drones. Dans ce chapitre, nous avons choisi de présenter uniquement les résultats bruts d'estimation de profondeur de l'algorithme. Une régularisation de ces cartes permettrait d'éliminer les valeurs aberrantes et d'améliorer ainsi le rendu visuel de la carte de profondeur. Cette question a été abordée de manière marginale au cours de ces travaux de thèse et les résultats de régularisation sont présentés dans la discussion de ce manuscrit.

2.4 Imageur chromatique

Dans cette section nous nous intéressons au potentiel d'estimation de profondeur d'un imageur chromatique à capacité 3D. L'utilisation d'un imageur chromatique nécessite le développement d'un nouvel algorithme d'estimation de profondeur. En effet, les scènes à l'origine des images de chaque canal RVB ne sont pas identiques mais simplement corrélées. Ainsi, les algorithmes traditionnels d'estimation de profondeur proposés dans la littérature en MA-DFD ne peuvent pas être utilisés. Nous présentons tout d'abord la généralisation de l'algorithme GLDFD dans un cas multi-images, dans lequel plusieurs acquisitions d'une même scène sont réalisées avec un détecteur noir et blanc. Cet algorithme, dit MA-GLDFD, est testé sur des images simulées. Puis nous développons une nouvelle version de l'algorithme dans le cas où l'imageur est chromatique, à l'aide d'une modélisation de la scène dans la représentation de la luminance et des chrominances. Cet algorithme, que nous appellerons SAC-GLDFD est tout d'abord testé sur des images simulées, puis sur les images produites par un prototype d'imageur chromatique, construit à partir d'un objectif chromatique conçu à l'IOGS. La précision d'estimation de profondeur est caractérisée sur l'axe par comparaison avec les valeurs d'un télémètre puis nous présentons des cartes de profondeurs comparées aux cartes produites par la Kinect.

2.4.1 Généralisation de l'algorithme mono-image au cas multi-image

Le formalisme utilisé pour définir le critère GL dans le cas de l'estimation de profondeur mono-image (SA-DFD) (équation (1.44) page 28) est suffisamment général pour pouvoir traiter le cas multi-image. Dans un premier temps, l'algorithme GLDFD est généralisé au cas où l'on dispose de deux ou plusieurs images de la même scène avec différentes configurations optiques (cas MA-DFD), puis le cas où le détecteur est couleur et où l'on dispose de trois images correspondant à l'acquisition de l'image suivant une bande spectrale différente (SAC-DFD). C'est notamment ce deuxième cas qui nous intéresse pour traiter les images produites par un imageur chromatique.

2.4.2 Généralisation de l'algorithme GLDFD au cas MA-DFD

2.4.2.1 Critère de sélection de FEP

Nous considérons tout d'abord le cas d'estimation de profondeur à partir de plusieurs images d'une même scène acquises dans des configurations différentes d'un même imageur. Les relations entre la scène et les deux images sont alors les suivantes :

$$\begin{aligned} \mathbf{y}_1 &= H_1 \mathbf{x} + \mathbf{b}_1 \\ \mathbf{y}_2 &= H_2 \mathbf{x} + \mathbf{b}_2, \end{aligned}$$

où H_i correspond à la matrice de convolution associée à la FEP de l'imageur dans la configuration i . Les expressions précédentes peuvent s'écrire sous la forme générale :

$$\mathbf{Y} = H_M \mathbf{x} + \mathbf{B}, \quad (2.6)$$

où :

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \mathbf{B} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} H_M = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix}.$$

Nous retrouvons ici le formalisme utilisé pour modéliser l'image dans la section 1.2.1 du chapitre 1, sauf qu'ici le vecteur de données est deux fois plus long dans le cas SA-DFD. Sous l'hypothèse d'un bruit indépendant et identiquement distribué gaussien, la vraisemblance des données s'écrit :

$$P(\mathbf{Y}|\mathbf{x}, \sigma_b^2) = (2\pi\sigma_b^2)^{-N} e^{-\frac{\|\mathbf{Y} - H_M \mathbf{x}\|^2}{2\sigma_b^2}} \quad (2.7)$$

Nous utilisons ensuite le même a priori sur la scène que dans le cas SA-DFD (équation (1.27)) et d'après la Proposition 1 de l'annexe A nous pouvons écrire :

$$L(\mathbf{Y}|\boldsymbol{\theta}) = \left| \frac{Q_y(\boldsymbol{\theta})}{2\pi} \right|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{Y}^t Q_y(\boldsymbol{\theta}) \mathbf{Y}\right). \quad (2.8)$$

avec

$$Q_y(\boldsymbol{\theta}) = \frac{1}{\sigma_b^2} (I - H_M (H_M^t H_M + \alpha D^t D)^{-1} H_M^t). \quad (2.9)$$

En notant $P_\psi = I - H_M (H_M^t H_M + \alpha D^t D)^{-1} H_M^t$ et $\psi \triangleq \{H_M, \alpha\}$, la vraisemblance (2.8) s'écrit alors :

$$L(\mathbf{Y}|\psi, \sigma_b^2) = (2\pi\sigma_b^2)^{-\left(\frac{2N-1}{2}\right)} |P_\psi|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{Y}^t \frac{P_\psi}{\sigma_b^2} \mathbf{Y}\right), \quad (2.10)$$

car à présent le vecteur \mathbf{Y} est de taille $2N$. La maximisation de la vraisemblance (2.10) par rapport au paramètre σ_b^2 implique :

$$L(\mathbf{Y}|\psi, \hat{\sigma}_b^2) \propto |P_\psi|_+^{\frac{1}{2}} (\mathbf{Y}^t P_\psi \mathbf{Y})^{-\left(\frac{2N-1}{2}\right)}.$$

Maximiser la vraisemblance revient alors à minimiser le critère GL_{MA} défini par :

$$\text{GL}_{\text{MA}}(\psi) = \mathbf{Y}^t P_\psi \mathbf{Y} |P_\psi|_+^{\frac{-1}{(2N-1)}} \quad (2.11)$$

Finalement, l'expression formelle du critère GL_{MA} est très proche de celle du critère GL dans l'équation (1.44) page 28, grâce au formalisme très général utilisé dans l'équation (2.6). La différence provient simplement de la construction du vecteur \mathbf{Y} et de la matrice H_M qui contiennent respectivement les différentes acquisitions et les différentes FEP associées aux modes d'acquisition. L'implémentation du critère GL_{MA} est identique à celle utilisée dans le cas SA-DFD et l'algorithme MA-GLDFD suit le même schéma que l'algorithme GLDFD présenté à la figure 1.2 à la page 29, sauf qu'il prend en entrée la concaténation d'une fenêtre de chaque image.

2.4.2.2 Validation sur des images simulées

Dans ce paragraphe, l'algorithme MA-GLDFD est validé sur des images simulées à partir de scènes respectant l'a priori gaussien puis sur des scènes naturelles. Les FEP d'un système optique sont simulées à partir de la méthode présentée en annexe B. Les caractéristiques optiques sont les mêmes que celles présentées dans le tableau 1.1 page 31. Les deux images sont acquises avec une modification de la mise au point, placée successivement à 2 m et à 3 m. L'effet de grandissement entre les deux images n'est pas modélisé ici. La figure 2.19 présente les résultats d'estimation de profondeur pour un ensemble de fenêtres de taille 21×21 pixels simulées à partir de de scènes respectant l'a priori gaussien, avec une valeur de α variant successivement 10^{-4} , 10^{-3} , 10^{-2} et 10^{-1} . L'estimation de profondeur est réalisée sur chaque fenêtre en utilisant le critère GL_{MA} .

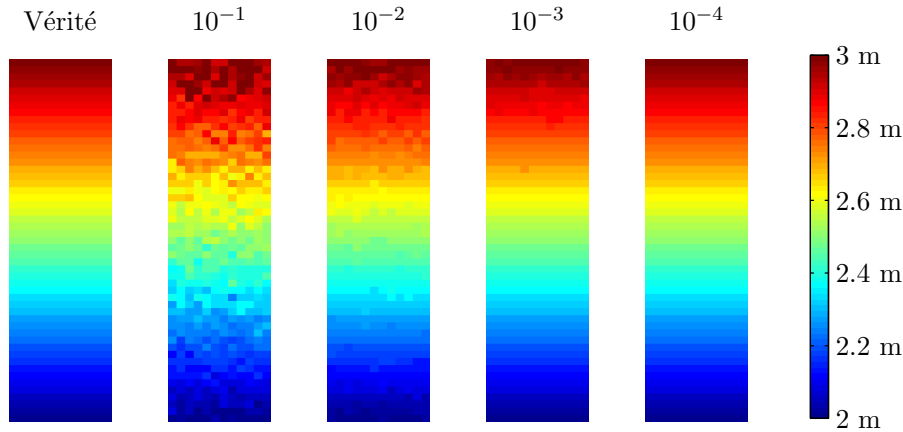


Fig. 2.19: Validation de l'algorithme MA-GLDFD sur des scènes vérifiant l'a priori gaussien, pour différentes valeurs du paramètre α , paramètre correspondant à l'inverse du rapport signal à bruit.

Les valeurs de profondeurs estimées sont très proches de la vérité, même avec un haut niveau de bruit. La comparaison des résultats de la figure 2.19 avec ceux obtenus dans le

cas mono-image (figure 1.4 page 32) montre que la variance des résultats est beaucoup plus faible dans le cas MA-DFD ce qui n'est pas surprenant car l'ajout d'une image supplémentaire facilite beaucoup l'estimation.

La figure 2.21 présente ensuite les résultats d'estimation de profondeur de l'algorithme MA-GLDFD sur des images simulées à partir d'extraits de scènes naturelles (a) à (d) présentées à la figure 2.20.

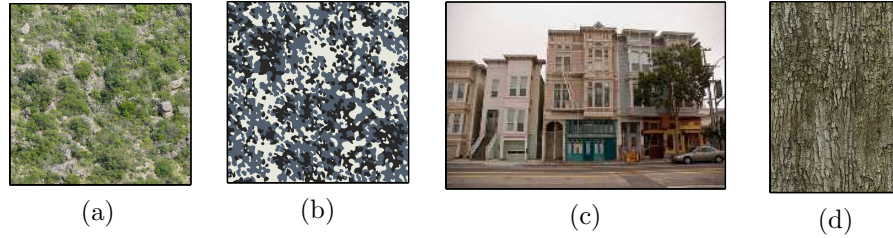


Fig. 2.20: Scènes utilisées pour valider l'algorithme MA-GLDFD sur des scènes naturelles.

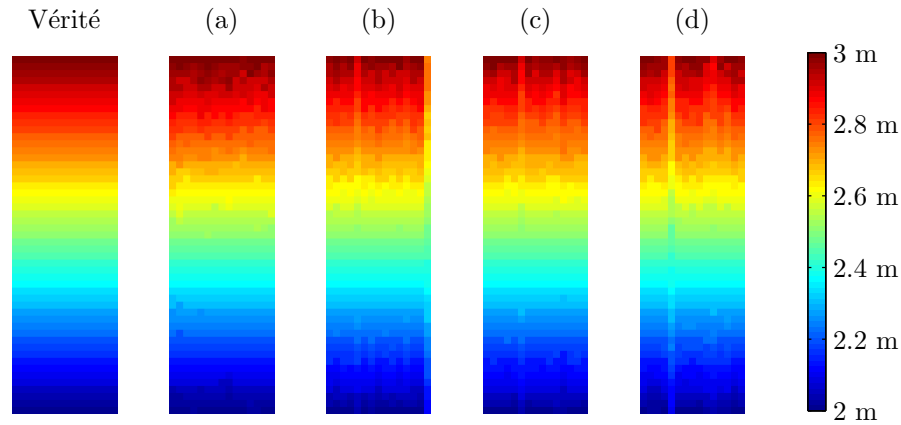


Fig. 2.21: Validation de l'algorithme MA-GLDFD sur les scènes naturelles (a) à (d) présentées à la figure 2.20.

Ces résultats montrent que l'algorithme MA-GLDFD fonctionne également dans le cas de scènes naturelles, ne respectant pas forcément l'a priori gaussien. Contrairement aux résultats de l'algorithme GLDFD, présentés à la figure 1.11 page 37, aucune valeur aberrante n'apparaît. L'utilisation de plusieurs images améliore donc nettement la robustesse de l'estimation de profondeur par DFD vis-à-vis d'une texture ne respectant pas l'a priori gaussien.

2.4.3 Généralisation de l'algorithme GLDFD au cas SAC-DFD

À présent nous considérons le cas de l'acquisition d'une seule scène avec un imageur chromatique et un détecteur couleur. Les images produites par les trois canaux RVB sont

alors modélisées de la manière suivante :

$$\mathbf{y}_R = H_R \mathbf{x}_R + \mathbf{b}_R \quad (2.12)$$

$$\mathbf{y}_V = H_V \mathbf{x}_V + \mathbf{b}_V \quad (2.13)$$

$$\mathbf{y}_B = H_B \mathbf{x}_B + \mathbf{b}_B, \quad (2.14)$$

où H_c correspond à la matrice de convolution associée à la FEP du canal c . Selon le type de détecteur, cette matrice prend ou non en compte le sous-échantillonnage lié à la présence d'un filtre de Bayer, comme discuté dans la section 1.5.3.

2.4.3.1 Modélisation de la scène

La difficulté ici est que les trois vecteurs \mathbf{x}_i sont partiellement corrélés. La figure 2.23 illustre ce phénomène en représentant pour les trois scènes de la figure 2.22, la répartition des pixels du canal R en fonction du canal V, du canal R en fonction du canal B et du canal B en fonction du canal V.

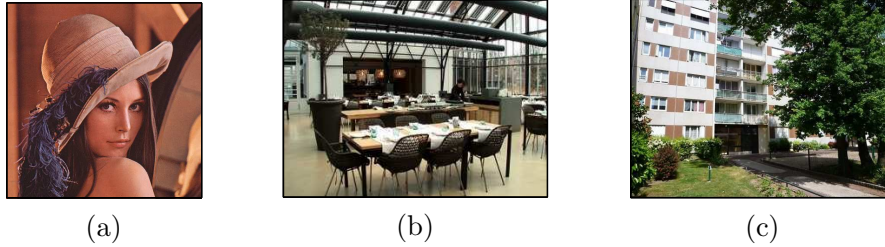


Fig. 2.22: Trois scènes naturelles

La corrélation entre les canaux RVB est visible sur les trois scènes, bien que son importance soit variable. Notamment, la scène (b) présentant beaucoup de régions en noir et blanc présente une corrélation plus importante des trois canaux RVB que les deux autres scènes, ce qui est prévisible car dans les images en noir et blanc les trois canaux sont identiques.

La corrélation partielle des trois composantes RVB rend délicate la définition d'une densité de probabilité jointe de ces trois composantes. Pour contourner ce point, nous proposons de décomposer la scène dans l'espace de la luminance et des chrominances rouge/verte et jaune/bleue (LC1C2). Les composantes de luminance et de chrominances peuvent en effet être considérées comme approximativement décorréliées selon plusieurs références, en particulier [Condat, 2009]. Cette décomposition s'écrit :

$$\begin{cases} \mathbf{x}_L = \frac{1}{\sqrt{3}}(\mathbf{x}_R + \mathbf{x}_V + \mathbf{x}_B) \\ \mathbf{x}_{C_1} = \frac{1}{\sqrt{2}}(-\mathbf{x}_R + \mathbf{x}_V) \\ \mathbf{x}_{C_2} = \frac{1}{\sqrt{6}}(-\mathbf{x}_R - \mathbf{x}_V + 2\mathbf{x}_B). \end{cases} \quad (2.15)$$

La figure 2.24 montre la répartition des pixels dans la décomposition LC1C2 pour les scènes de la figure 2.22.

Cette figure illustre la diminution de la corrélation entre les composantes avec la décomposition LC1C2 au lieu de la décomposition RVB. Avec cette décomposition, les trois

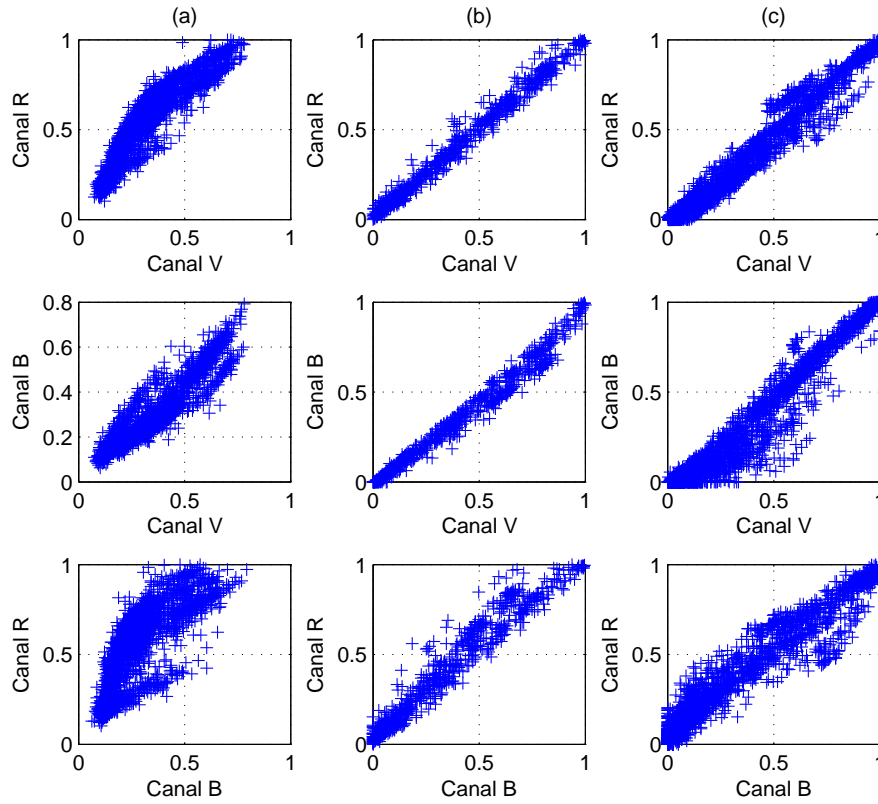


Fig. 2.23: Répartition des pixels R/V R/B et B/V pour les images présentées à la figure 2.22.

scènes peuvent s'écrire :

$$\begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_V \\ \mathbf{x}_B \end{bmatrix} = M \begin{bmatrix} \mathbf{x}_L \\ \mathbf{x}_{C_1} \\ \mathbf{x}_{C_2} \end{bmatrix} = \begin{bmatrix} a\mathbf{x}_L + b\mathbf{x}_{C_1} + c\mathbf{x}_{C_2} \\ d\mathbf{x}_L + e\mathbf{x}_{C_1} + f\mathbf{x}_{C_2} \\ g\mathbf{x}_L + h\mathbf{x}_{C_1} + i\mathbf{x}_{C_2} \end{bmatrix} \quad (2.16)$$

où :

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \quad (2.17)$$

Nous modélisons ensuite les distributions des trois composantes luminance et chrominances en utilisant un a priori gaussien :

$$P(\mathbf{x}_L, \sigma_{x_L}^2) \propto \exp\left(-\frac{\|D\mathbf{x}_L\|^2}{2\sigma_{x_L}^2}\right) \quad (2.18)$$

$$P(\mathbf{x}_{C_1}, \sigma_{x_C}^2) \propto \exp\left(-\frac{\|D\mathbf{x}_{C_1}\|^2}{2\sigma_{x_C}^2}\right) \quad (2.19)$$

$$P(\mathbf{x}_{C_2}, \sigma_{x_C}^2) \propto \exp\left(-\frac{\|D\mathbf{x}_{C_2}\|^2}{2\sigma_{x_C}^2}\right), \quad (2.20)$$

où D est la concaténation des matrices de convolution associées aux dérivées horizontales et verticales d'ordre 1. En notant $\mathbf{X}^{LC_1C_2} = [\mathbf{x}_L^t \mathbf{x}_{C_1}^t \mathbf{x}_{C_2}^t]^t$ et en considérant que les trois

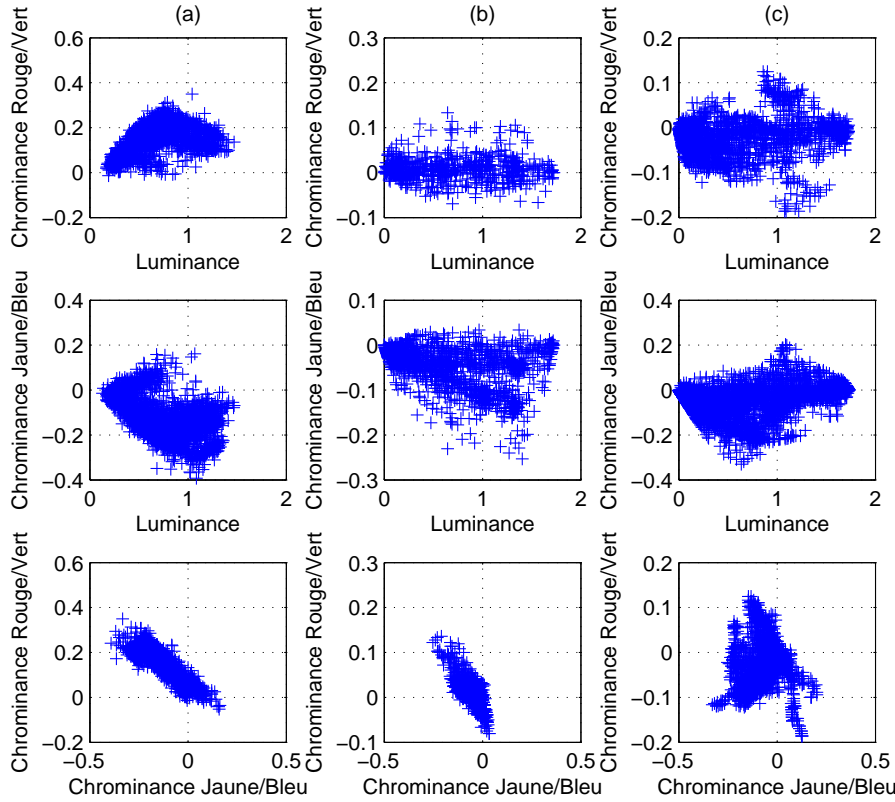


Fig. 2.24: Répartition des pixels L/C_1 , L/C_2 et C_1/C_2 pour les images présentées à la figure 2.22.

composantes sont indépendantes, la densité de probabilité de la scène est modélisée de la manière suivante :

$$P(\mathbf{X}^{LC_1C_2}, \sigma_{x_C}^2, \mu_c) \propto \exp\left(-\frac{\|D_C(\mu_c)\mathbf{X}\|^2}{2\sigma_{x_C}^2}\right) \quad (2.21)$$

Avec :

$$D_C = \begin{bmatrix} \sqrt{\mu_c}D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{bmatrix}. \quad (2.22)$$

Comme dans [Condat, 2009], le paramètre μ_c est introduit dans l'équation (2.21) afin de modéliser le rapport entre les variances des gradients des chrominances $\sigma_{x_C}^2$ et la variance des gradients de la luminance $\sigma_{x_L}^2$. Ce paramètre varie entre 0 et 1 et ses deux bornes correspondent à deux types de scènes différentes. Lorsque μ_c tend vers 0, cela signifie que la variance de la luminance est très élevée par rapport à celle des chrominances. Or, une faible chrominance correspond à une scène en noir et blanc. Ainsi, le cas μ_c tend vers 0 correspond à considérer que la scène est noire et blanche. Lorsque μ_c tend vers 1, cela revient à négliger la corrélation entre les trois canaux RVB [Condat, 2009]. Ce cas revient à considérer que les trois images sont produites à partir de scènes décorréélées, donc que l'objet a une texture décorréélées suivant les trois spectres RVB. Ce cas est plutôt un cas théorique dans notre application.

2.4.3.2 Critère de sélection de FEP

Remplacer (2.16) dans (2.14) permet d'écrire :

$$\mathbf{y}_R = H_R(a\mathbf{x}_L + b\mathbf{x}_{C_1} + c\mathbf{x}_{C_2}) + \mathbf{b}_R \quad (2.23)$$

$$\mathbf{y}_V = H_V(d\mathbf{x}_L + e\mathbf{x}_{C_1} + f\mathbf{x}_{C_2}) + \mathbf{b}_V \quad (2.24)$$

$$\mathbf{y}_B = H_B(g\mathbf{x}_L + h\mathbf{x}_{C_1} + i\mathbf{x}_{C_2}) + \mathbf{b}_B, \quad (2.25)$$

Cette expression se simplifie en utilisant les notations :

$$\mathbf{X}^{LC_1C_2} = \begin{bmatrix} \mathbf{x}_L \\ \mathbf{x}_{C_1} \\ \mathbf{x}_{C_2} \end{bmatrix} \mathbf{Y} = \begin{bmatrix} \mathbf{y}_R \\ \mathbf{y}_V \\ \mathbf{y}_B \end{bmatrix} \mathbf{B} = \begin{bmatrix} \mathbf{b}_R \\ \mathbf{b}_V \\ \mathbf{b}_B \end{bmatrix}. \quad (2.26)$$

L'équation (2.25) devient alors :

$$\mathbf{Y} = H_C \mathbf{X}^{LC_1C_2} + \mathbf{B}, \quad (2.27)$$

où

$$H_C = \begin{bmatrix} aH_R & bH_R & cH_R \\ dH_V & eH_V & fH_V \\ gH_B & hH_B & iH_B \end{bmatrix}. \quad (2.28)$$

La Proposition 1 de l'annexe A s'applique également ici, ce qui permet d'écrire :

$$L(\mathbf{Y}|\boldsymbol{\theta}) = \left| \frac{Q_Y(\boldsymbol{\theta})}{2\pi} \right|_+^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{Y}^t Q_Y(\boldsymbol{\theta}) \mathbf{Y} \right). \quad (2.29)$$

avec

$$Q_Y(\boldsymbol{\theta}) = \frac{1}{\sigma_b^2} (I - H_C(H_C^t H_C + \alpha D_C^t D_C)^{-1} H_C^t), \quad (2.30)$$

$\boldsymbol{\theta}$ est alors défini par $\{H_C, \sigma_b^2, \alpha\}$. Ici le déterminant de Q_Y a trois valeurs propres nulles, qui sont liées chacune aux trois composantes $\mathbf{x}_L, \mathbf{x}_{C_1}, \mathbf{x}_{C_2}$. En notant $\psi \triangleq \{H_C, \alpha\}$ et $P_\psi = I - H_C(H_C^t H_C + \alpha D_C^t D_C)^{-1} H_C^t$, la vraisemblance (2.29) s'écrit alors :

$$L(\mathbf{Y}|\psi, \sigma_b^2) = (2\pi\sigma_b^2)^{-\left(\frac{3N-3}{2}\right)} |P_\psi|_+^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{Y}^t \frac{P_\psi}{\sigma_b^2} \mathbf{Y} \right). \quad (2.31)$$

En maximisant la vraisemblance donnée par l'équation (2.31) par rapport à σ_b^2 nous obtenons alors la vraisemblance généralisée :

$$L(\mathbf{Y}|\psi, \hat{\sigma}_b^2) \propto |P_\psi|_+^{\frac{1}{2}} (\mathbf{Y}^t P_\psi \mathbf{Y})^{-\left(\frac{3N-3}{2}\right)}.$$

Maximiser la vraisemblance revient donc à minimiser le critère GL-SAC défini par :

$$\text{GL}_{\text{SAC}}(\psi) = \mathbf{Y}^t P_\psi \mathbf{Y} |P_\psi|_+^{\frac{-1}{(3N-3)}}. \quad (2.32)$$

L'implémentation du critère GL_{SAC} est identique à celle utilisée dans le cas SA-DFD, et l'algorithme SAC-GLDFD suit le même schéma que l'algorithme GLDFD présenté à la figure 1.2 à la page 29, sauf qu'il prend en entrée la concaténation d'une fenêtre de chaque image RVB.

2.4.3.3 Validation sur des images simulées

Dans cette simulation, un imageur idéal faiblement chromatique est modélisé à partir des paramètres du tableau 2.2. Les FEP des trois canaux RVB sont simulées à partir de la méthode présentée en annexe B et, pour simplifier, les FEP des canaux R,V et B sont simulées respectivement pour les longueurs d'onde 640, 550 et 470 nm, qui sont les longueurs d'onde correspondant typiquement aux maxima des filtres R,V et B d'un détecteur couleur.

Focales (RVB)	25 mm
Ouverture $F/\#$	2.5
p_x	5 μm
p_0 (RVB)	1.9 2 et 2.1 m

Tableau 2.2: Caractéristiques du système optique idéal utilisé dans les simulations.

Dans un premier temps, l'algorithme SAC-GLDFD est testé sur des images simulées à partir de scènes vérifiant l'a priori (2.21). Ces scènes sont générées dans l'espace de la luminance et des chrominances en utilisant la méthode décrite dans l'annexe C pour une valeur de μ_c fixée à 0.05 et $\sigma_{x_C}=50$. Puis elles sont converties dans la représentation RVB et chaque canal est convolué à la FEP_c associée à une profondeur. La figure 2.25 présente les résultats d'estimation de profondeur pour un ensemble de fenêtres générées suivant la même procédure que dans le paragraphe 1.3.1.1 avec une valeur de α variant successivement entre 10^{-4} , 10^{-3} , 10^{-2} et 10^{-1} . La taille des fenêtres est ici de 12×12 pixels. Les

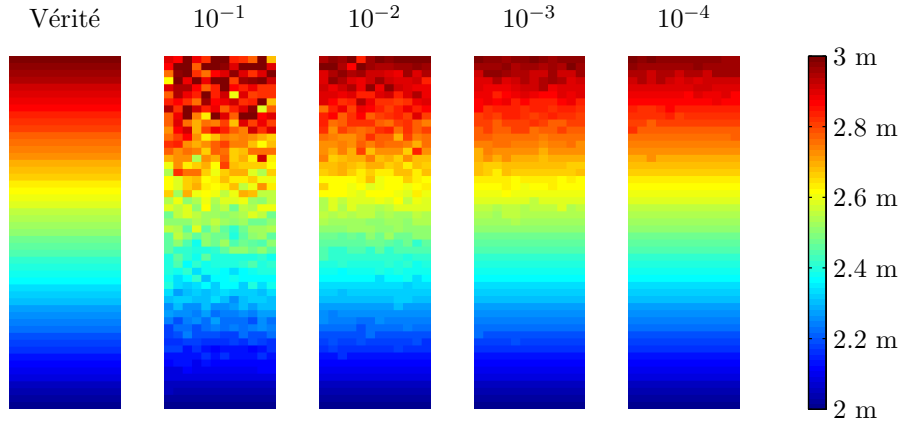


Fig. 2.25: Estimation de profondeur avec le critère GL-SAC dans le cas de scènes vérifiant l'a priori gaussien, pour différentes valeurs de α .

résultats d'estimations de profondeur sont cohérents avec la vérité et la variance des résultats diminue lorsque le rapport signal à bruit s'améliore. Dans le cas précédent, les scènes sont générées avec la même valeur de μ_c que celle utilisée dans l'algorithme. Or, en pratique, la valeur de μ_c n'est pas définie. La figure 2.26 montre les résultats d'estimation de profondeur en utilisant cette valeur de paramètre pour un quinzaine de fenêtres extraites de quatre scènes naturelles présentées à la figure 2.20. Les fenêtres sont générées suivant la même procédure que dans le cas MA-DFD présenté précédemment, sauf que les images des canaux RVB sont simulées à partir des canaux RVB de la scène d'origine, convolués aux FEP qui leur sont associées avec l'ajout d'un bruit blanc gaussien d'écart-type 0.01.

Nous fixons dans un premier temps, comme Condat, μ_c à 0.05 [Condat, 2009]. La figure 2.26

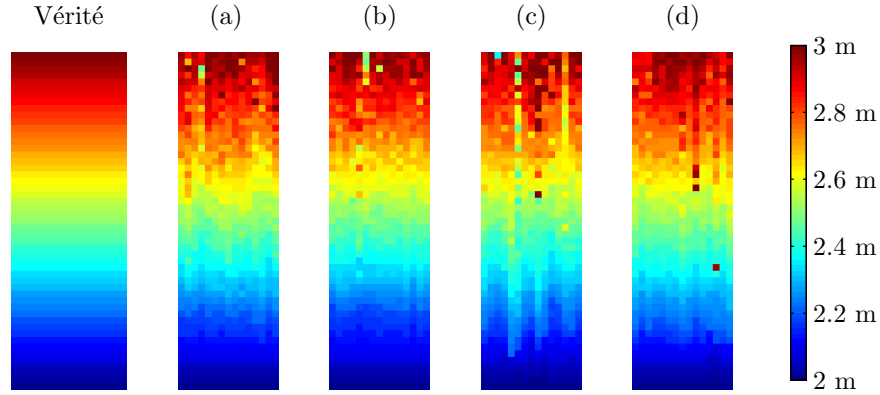


Fig. 2.26: Résultats de l'estimation de la profondeur avec l'algorithme SAC-GLDFD. Les scènes RVB sont des scènes naturelles.

montre que l'algorithme SAC-GLDFD estime correctement la profondeur pour des scènes naturelles, avec cette valeur de paramètre. Afin d'étudier plus précisément l'effet du choix du paramètre μ_c , le test précédent est reproduit pour 20 scènes naturelles, dont 10 d'entre elles sont présentées à l'annexe F, et pour différentes valeurs de μ_c . Le tableau 2.3 présente l'écart moyen en valeur absolue à la valeur vraie sur toutes les profondeurs et toutes les scènes, obtenu pour chaque valeur de μ_c . Nous pouvons constater que la valeur $\mu_c = 0.05$ a

μ_c	0.0005	0.005	0.05	0.5
Écart moyen à la vérité en cm	10.7	5.9	4.6	6.1

Tableau 2.3: Écart moyen à la vérité et écart-type obtenus sur un ensemble d'images simulées pour différentes valeurs de μ_c .

les meilleures performances d'estimation de profondeur sur ces scènes naturelles, ce qui est cohérent avec la valeur choisie par [Condat, 2009].

Par ailleurs, nous pouvons nous demander quel sera le comportement de l'algorithme SAC-GLDFD dans le cas de scènes en niveau de gris. La simulation précédente est donc reproduite pour les mêmes scènes, mais ici les images des canaux RVB sont simulées à partir de la conversion en niveau de gris des canaux RVB de la scène d'origine. L'objectif est d'étudier la robustesse de l'algorithme SAC-GLDFD lorsque la valeur de μ_c fixée dans l'algorithme est très différente de la valeur réelle (ici $\mu_c = 0$). La figure 2.27 présente les résultats obtenus sur quatre scènes avec $\mu_c = 0.05$.

Même pour des scènes en niveau de gris, l'algorithme SAC-GLDFD estime correctement la profondeur. En conclusion, l'algorithme SAC-GLDFD permet d'estimer la profondeur d'une grande variété de scènes, même si la valeur de μ_c est fixée a priori.

2.4.4 Réalisation expérimentale

2.4.4.1 Conception d'un imageur chromatique

Un objectif chromatique a été conçu par J. Sabater, L. Jacubowicz et T. Avignon de l'IOGS dans le cadre d'un projet pédagogique visant à reproduire l'article [Guichard *et al.*,

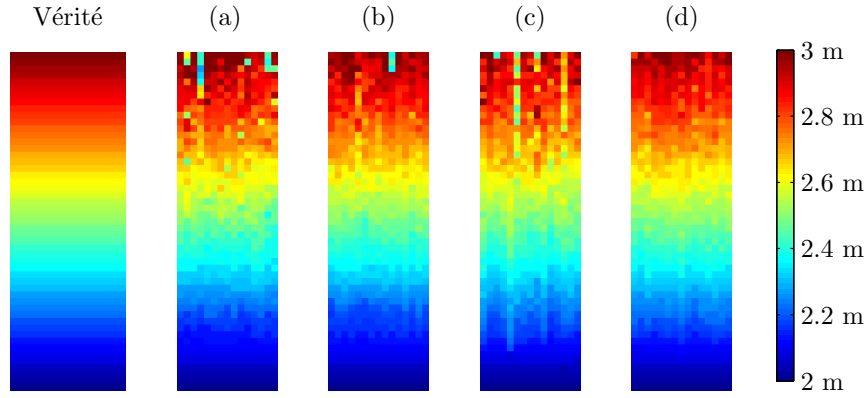


Fig. 2.27: Test de l'estimation de la profondeur. Les scènes sont naturelles et converties en niveau de gris

2009]. Cet objectif a pu être réalisé dans le cadre d'une collaboration IOGS-ONERA et nous avons ainsi pu tester l'algorithme SAC-GLDFD sur des images réelles. La figure 2.28 présente une photographie de l'imageur conçu ainsi que les plans des différentes lentilles. La distance focale de cet imageur est de 25 mm et l'ouverture de 4. Le chromatisme longitudinal, à savoir la différence entre les distances focales du canal rouge et du canal bleu est de $200\ \mu\text{m}$. Cet objectif a été conçu afin d'avoir les trois plans de mise au point RVB compris entre 1 et 5 m et d'avoir un chromatisme latéral nul. Ceci permet de ne pas avoir à corriger les effets de grandissement entre les images des trois canaux RVB. L'objectif est combiné à un détecteur couleur de la marque Stingray de taille 2056×2452 pixels comportant un filtre de Bayer, avec des pixels de taille $3.45\ \mu\text{m}$. Le champ total de l'imageur est alors d'environ (20° , 16°). La mise au point de l'imageur est réglée afin d'avoir un plan de mise au point du canal bleu à 1.9 m, du canal vert à 2.7 m et du canal rouge à 4.5 m.

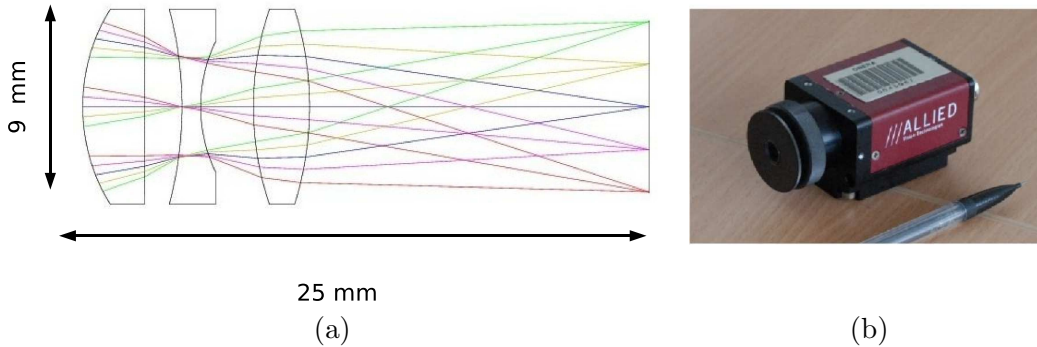


Fig. 2.28: (a) Plan des lentilles constituant l'objectif chromatique. (b) Objectif chromatique combiné au détecteur Stingray.

2.4.4.2 Calibrage des FEP sur l'axe

Les FEP de cet imageur sont calibrées sur l'axe en utilisant la méthode de Delbracio, décrite à la section 2.2.1.3, entre 1 et 4.5 m avec un pas de 5 cm. La figure 2.29 présente

plusieurs FEP calibrées pour les trois canaux à trois profondeurs différentes. L'effet de l'aberration chromatique entre les canaux est nettement visible sur les FEP.

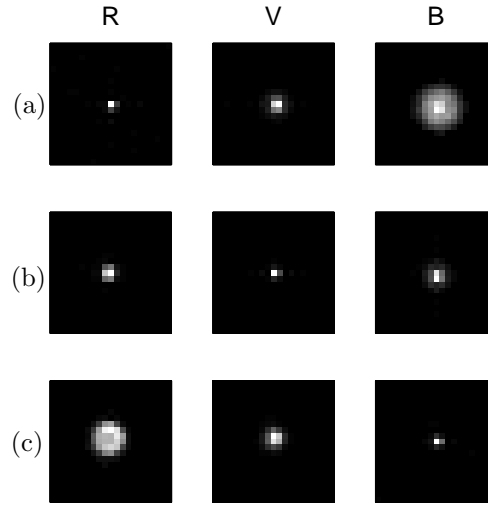


Fig. 2.29: FEP calibrées par la méthode de [Delbracio *et al.*, 2011] pour trois profondeurs : (a) 4.7 m, (b) 2.7 m et (c) 1.9 m.

2.4.4.3 Calibrage des FEP hors axe

Le calibrage des FEP est également effectué hors axe afin de mesurer qualitativement la variation de la FEP dans le champ. L'objectif est de déterminer la région de l'espace autour de l'axe optique pour laquelle la FEP peut être considérée comme constante. La figure 2.30 présente plusieurs FEP acquises pour différentes valeurs du champ. Elle montre que

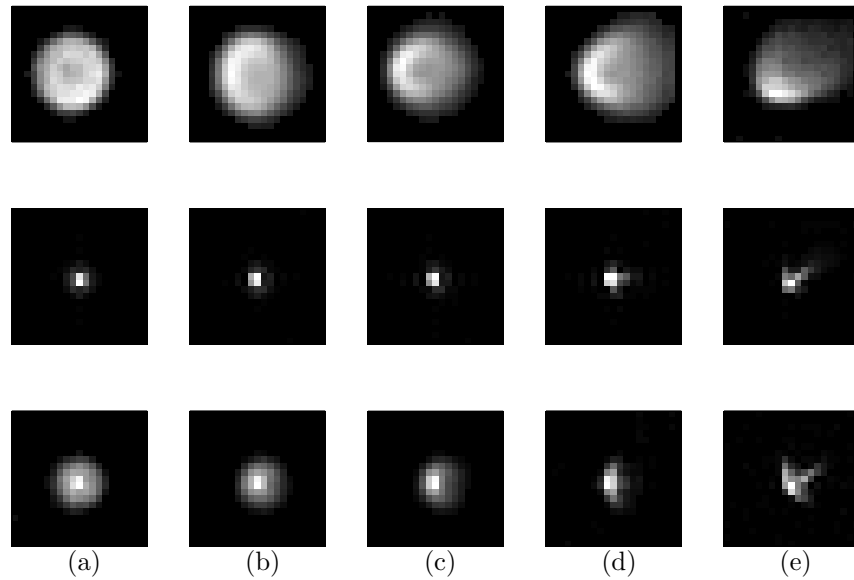


Fig. 2.30: Variation de la FEP pour différentes positions angulaires de la mire (θ_h, θ_v) par rapport à l'axe optique, pour trois profondeurs (de haut en bas 1 m, 2.7 m et 5 m). (a) sur l'axe (b) $(0^\circ, 1^\circ)$, (c) $(0^\circ, 3^\circ)$ et (d) $(0^\circ, 5^\circ)$ (e) $(4^\circ, 5^\circ)$. Le coin en haut à droite de l'image correspond aux angles $(10^\circ, 8^\circ)$.

la variation de la FEP est plus importante dans le champ que pour l'objectif du commerce utilisé avec l'appareil photographique Nikon. Dans la suite, la FEP sera supposée invariante dans un angle de 1° de champ. La construction de carte de profondeur nécessite donc le calibrage des FEP dans le champ. L'image est donc découpée en 9 régions et la FEP est calibrée au milieu de chacune. La figure 2.31 montre la découpe des 9 régions sur l'image et des exemples de FEP calibrées dans chaque région pour les trois canaux à 3.9 m.

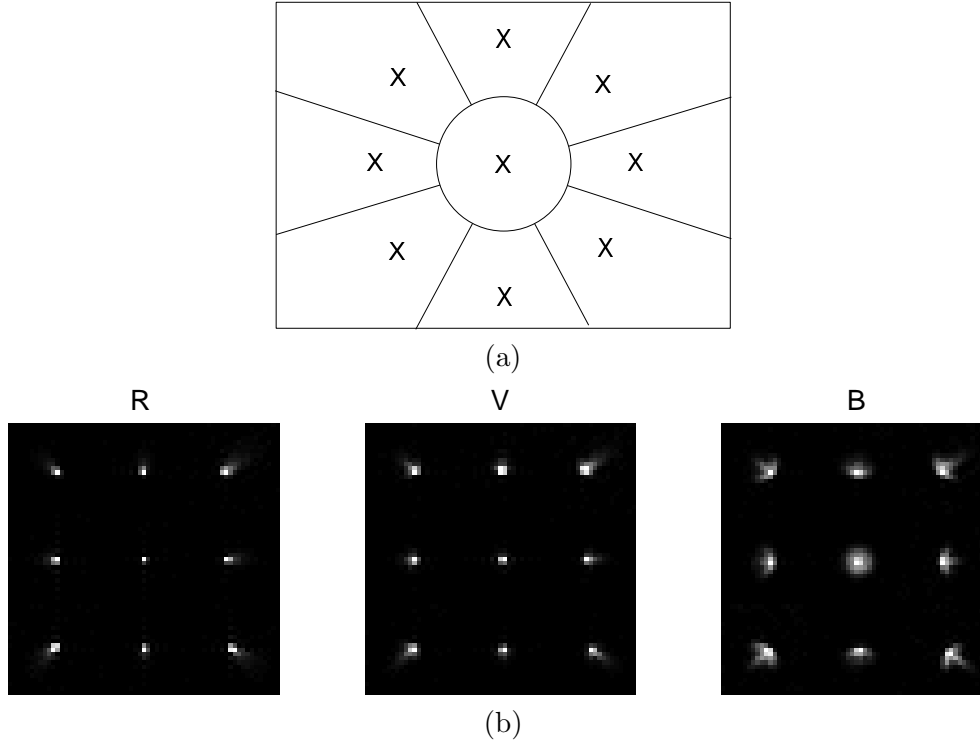


Fig. 2.31: (a) Découpe de l'image en 9 régions où la FEP est supposée constante.
(b) FEP calibrées dans les 9 régions de l'image pour les trois canaux à 3.9 m.

2.4.5 Mesures de distance sur l'axe

Dans un premier temps, la précision d'estimation de profondeur de l'imager chromatique est évaluée sur l'axe pour un ensemble de cibles. L'objectif est d'évaluer la robustesse de la méthode proposée vis-à-vis de la variation de texture mais également vis-à-vis de la variation de la FEP avec le spectre de la cible.

Scènes colorées

La figure 2.32 présente les résultats d'estimation de profondeur obtenus sur quatre des scènes colorées présentées à la figure 2.6. La profondeur est estimée à l'intérieur d'une région centrée de l'image brute de largeur 240×240 pixels qui correspond à une position angulaire de $(2^\circ, 2^\circ)$ dans les directions horizontale et verticale par rapport à l'axe optique. L'erreur liée au tilt est alors inférieure à 1 mm. Les images RVB sont extraites d'une fenêtre de taille 42×42 pixels sur l'image brute donc une fenêtre de taille 21×21 pixels est extraite de chaque canal. Ici, les données sont traitées comme si le détecteur était de type tri-CCD, donc l'un des canaux verts n'est pas utilisé et les trois canaux restant RV_1B sont traités sans modéliser le sous-échantillonnage. Les FEP de la famille sont celles calibrées sur chaque

canal sans faire d'interpolation. Le recouvrement étant de 50%, le calcul de la moyenne et de l'écart-type est effectué sur 100 résultats d'estimation de profondeur, dont 5% sont éliminés.

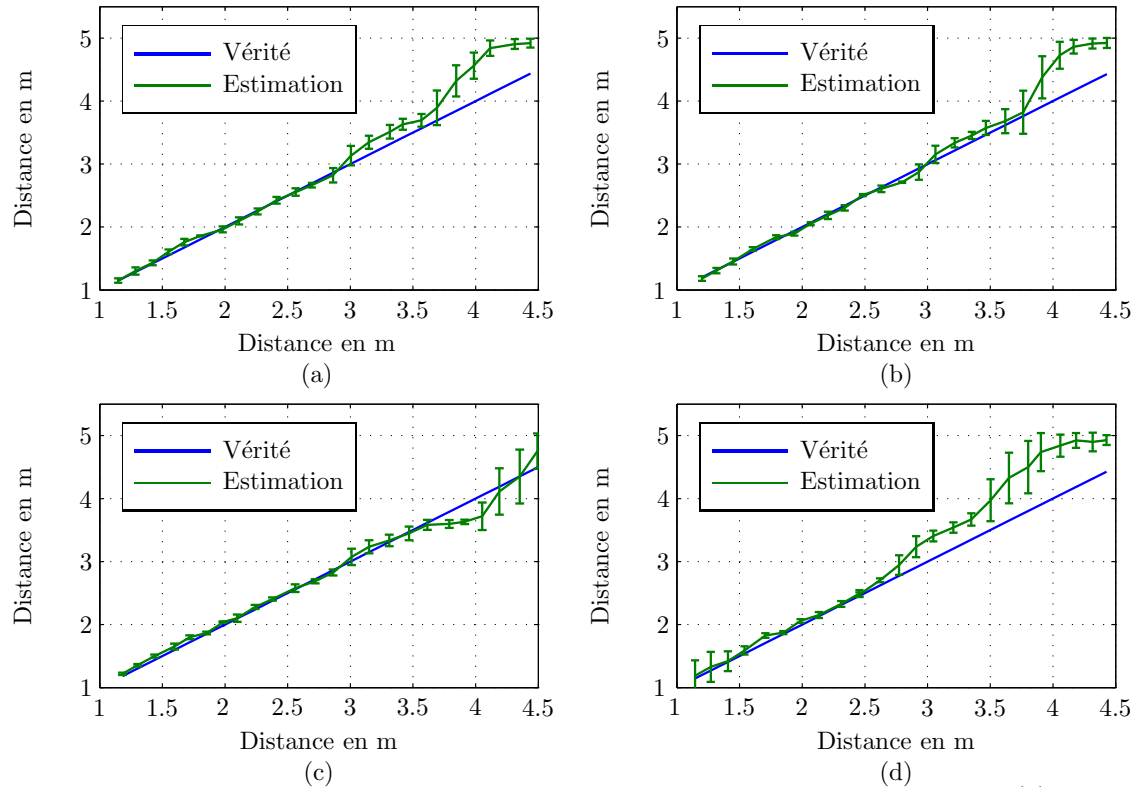


Fig. 2.32: Résultats d'estimation de profondeur sur quatre cibles colorées. (a) Cible « Fleurs », (b) Cible « Parterre », (c) Cible « Carreaux », (d) Cible « Ville ».

Pour les quatre cibles, l'estimation de profondeur a un biais négligeable et un écart-type inférieur à 5 cm entre 1 et 3 m. Ensuite le biais reste négligeable jusqu'à 4 m pour les cibles « Fleurs » et « Carreaux » (figures (a) et (c)) alors qu'il augmente pour les deux autres cibles. Bien que ces quatre cibles présentent différents spectres, les performances d'estimation de profondeur sont d'allure similaires entre 1 et 3 m. Sur cet exemple, la solution chromatique apparaît nettement plus robuste à la variation de la texture de la scène que l'imageur à pupille codée.

Scènes en niveau de gris

La série d'acquisitions précédente est reproduite en imprimant les mêmes cibles mais cette fois-ci en niveau de gris. La figure 2.33 présente les résultats d'estimation de profondeur ainsi obtenus. Les performances d'estimation de profondeur sont très similaires à celles obtenues lorsque les cibles sont en couleur, ce qui est cohérent avec l'étude réalisée sur les images simulées dans la section 2.4.3.3 qui a montré que l'algorithme SAC-GLDFD fonctionne pour des images couleurs et des images en noir et blanc.

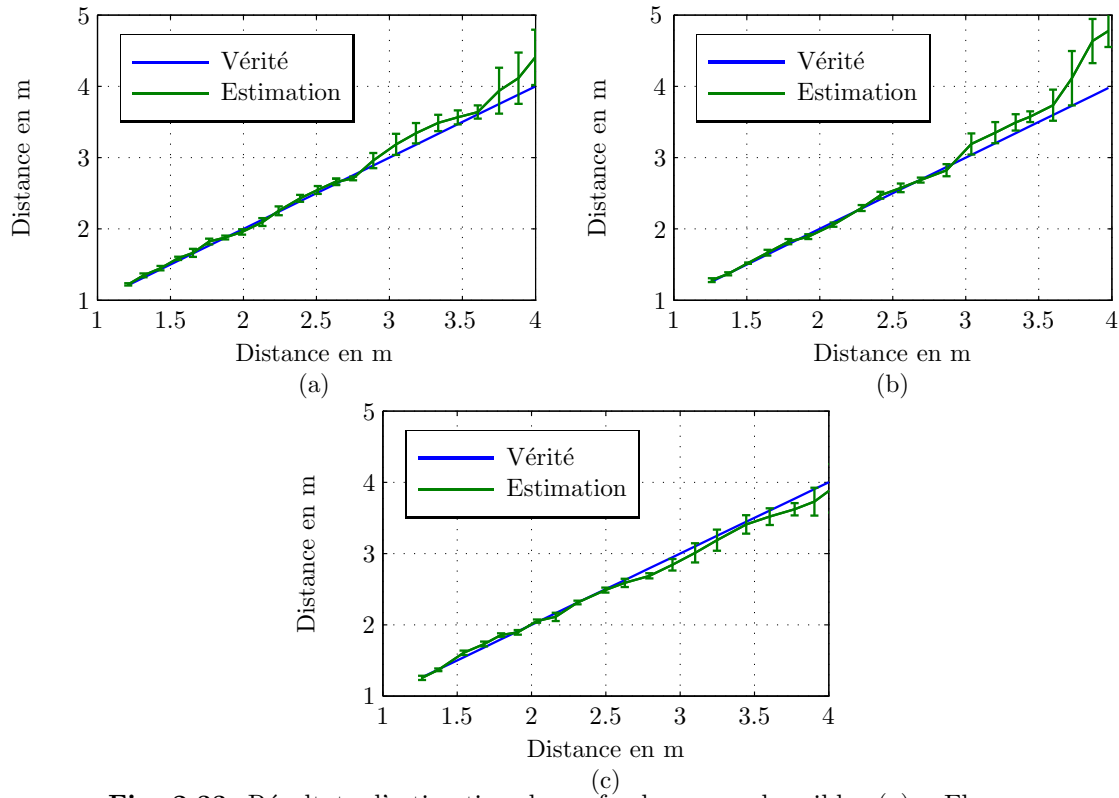


Fig. 2.33: Résultats d'estimation de profondeur pour les cibles (a) « Fleurs ». (b) Cible « Parterre ». (c) Cible « Carreaux » imprimées en noir et blanc.

Utilisation de l'algorithme MA-GLDFD sur des images couleurs issues d'un imageur chromatique

La figure 2.34 présente les résultats d'estimation de profondeur obtenus sur la cible « Fleurs » imprimée en couleur en utilisant l'algorithme MA-GLDFD qui repose sur l'hypothèse d'une scène unique pour les trois canaux et qui est donc adapté au cas de scènes en noir et blanc. L'estimation de profondeur fonctionne en moyenne mais, en comparant

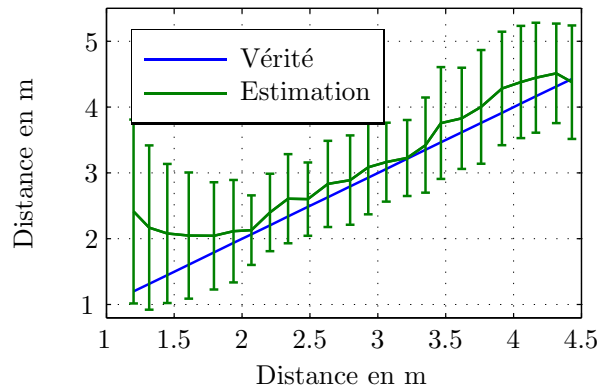


Fig. 2.34: Résultat d'estimation de profondeur en utilisant l'algorithme MA-GLDFD sur la cible « Fleurs » de la figure 2.6(b) imprimée en couleur.

avec la figure 2.32 (b), l'écart-type est beaucoup plus important qu'avec l'algorithme SAC-GLDFD. Ainsi, si l'algorithme SAC-GLDFD peut être utilisé à la fois pour des scènes en couleur et en noir et blanc, l'algorithme MA-GLDFD ne fonctionne pas sur des scènes en couleur.

Bilan

Ces différents tests montrent que pour une variété de cibles en couleur ou en noir et blanc, l'algorithme SAC-GLDFD et l'imageur chromatique permettent d'estimer la profondeur avec une précision importante. En effet, l'écart-type empirique est du même ordre de grandeur que le pas de calibrage, qui est de 5 cm, pour des distances comprises entre 1 et 3 m, voire 3.5 m pour certaines cibles. Ainsi, la méthode proposée est robuste à la variation de la texture et de la couleur de la scène. Par ailleurs, nos essais montrent qu'une étude plus approfondie pourrait être menée sur une plus grande diversité de cible et en utilisant un pas de calibrage plus fin, car l'écart-type mesuré est proche du pas de calibrage actuel.

2.4.6 Cartes de profondeur

À partir des FEP calibrées sur l'axe et hors d'axe, l'estimation de profondeur est réalisée sur plusieurs images composées de différents niveaux de profondeurs. Trois exemples de résultats sont présentés à la figure 2.35.

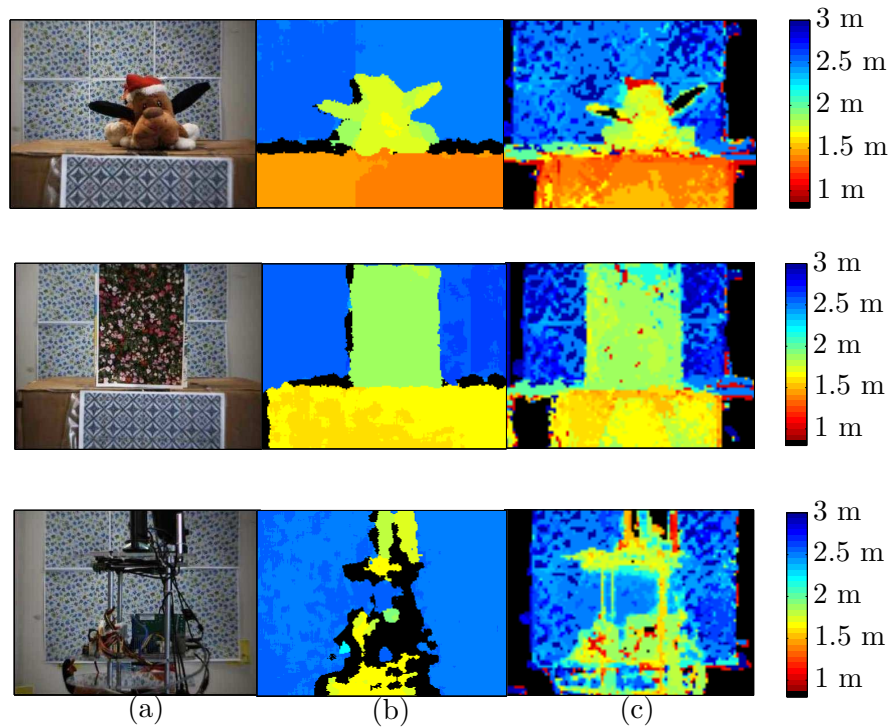


Fig. 2.35: (a) Acquisitions avec l'imageur chromatique. (b) Cartes de profondeur obtenues avec la Kinect. (c) Cartes de profondeur brutes obtenues avec l'algorithme SAC-GLDFD.

Les valeurs de profondeur estimées par l'algorithme SAC-GLDFD sont cohérentes avec celles de la Kinect, bien que plus bruitées. La dernière acquisition illustre le cas d'une scène pour laquelle la Kinect renvoie des valeurs aberrantes à cause de surfaces réfléchissantes et

des problèmes d'occultation sur les tubes de l'objet. L'imageur chromatique estime correctement la profondeur de cette scène et renvoie une carte plus bruitée mais plus complète. Le découpage de l'image en 9 zones de FEP constante permet de tenir compte de la variation de la FEP dans le champ mais le découpage reste encore trop grossier. Cela introduit des discontinuités sur la profondeur qui sont visibles à certains endroits des cartes de profondeur.

2.4.7 Discussion

Dans ce chapitre, nous avons développé et présenté les premiers résultats d'estimation de profondeur à partir d'un imageur à capacité 3D utilisant une optique chromatique. Sur les cibles couleurs et noir et blanc, l'estimation de profondeur a un biais négligeable entre 1 et 3 m et l'écart-type est inférieur à 5 cm à 1 m et de 10 cm à 3 m. Une étude plus précise devrait être menée, pour caractériser la précision d'estimation de profondeur car l'écart-type est souvent inférieur à 5 cm qui est le pas de calibrage.

Les cartes de profondeur obtenues apparaissent cohérentes avec celles obtenues par le système actif. La variation importante de la FEP dans le champ nécessite un calibrage hors axe et le découpage de l'image en régions de FEP constante introduit certaines discontinuités dans la carte de profondeur.

Nous présentons à présent deux points liés à l'utilisation d'un imageur chromatique qui à l'issue de ces travaux mériteraient une étude approfondie.

2.4.7.1 Variation de la FEP suivant le spectre de la scène

Le premier point concerne le calibrage des FEP chromatiques. Dans ce chapitre, les FEP sont calibrées sur des mires noir et blanc avec une lumière naturelle. Les FEP des canaux R, V et B correspondent donc à la FEP polychromatique pour toutes les longueurs d'onde appartenant respectivement aux spectres des filtres rouge, vert et bleu du détecteur. Or si la scène contient un spectre plus étroit, la FEP risque d'être légèrement différente. La figure 2.36 illustre ce phénomène. La figure 2.36(a) présente la FEP polychromatique obtenue à l'aide du logiciel Zemax pour un point source situé à 2 m de l'imageur chromatique. Les longueurs d'ondes utilisées pour simuler cette FEP appartiennent respectivement aux spectres des filtres rouge, vert et bleu d'une caméra couleur. La figure (b) correspond aux FEP monochromatiques pour la longueur d'onde au centre du spectre de chaque filtre.

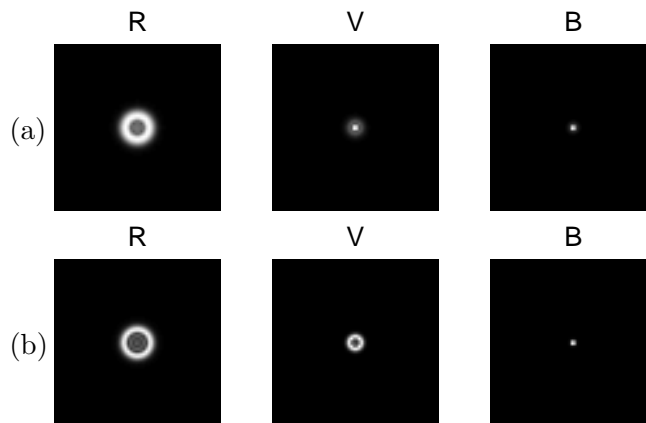


Fig. 2.36: (a) Simulation de FEP polychromatiques associées aux spectres complet des canaux RVB. (b) Simulation de FEP monochromatique respectivement à $0.64 \mu\text{m}$, $0.55 \mu\text{m}$ et $0.48 \mu\text{m}$.

Effectivement les FEP varient légèrement suivant le spectre, ce qui risque de perturber l'estimation de profondeur. Cependant nos tests sur plusieurs types de scènes avec des spectres de couleur différents conduisent à des performances similaires, ce qui indique que l'effet illustré à la figure 2.36 est ici négligeable. Cependant, il serait intéressant d'effectuer des tests supplémentaires sur des cibles texturées mais monochromatiques pour voir si les performances d'estimation de profondeur se dégradent.

2.4.7.2 Restauration de l'image RVB

Le deuxième point négligé dans l'analyse menée dans ce chapitre concerne la qualité des images RVB produites par l'imageur. En effet, l'accentuation du chromatisme, ou simplement sa non correction, réduit la qualité de l'image couleur car chaque canal possède son propre plan de mise au point. Pour restaurer l'image nette, une piste de travail est d'utiliser la méthode d'extension de profondeur de champ proposée dans [Guichard *et al.*, 2009]. Le principe est de transférer les informations hautes fréquences du canal le plus net vers les canaux les plus flous. Pour cela, les auteurs déterminent le canal le plus net par un critère de netteté de la forme :

$$M_i(m_0, n_0) = \frac{|\nabla y_i(m, n) - \nabla y_i(m_0, n_0)|}{\nabla y_i(m_0, n_0)}, \quad (2.33)$$

où i renvoie au canal R, V ou B. Le transfert de fréquence est réalisé par une opération du type :

$$y_{i,\text{out}} = y_{i,\text{in}} + a_{i,R} \text{HP}_{i,R}(y_R) + a_{i,V} \text{HP}_{i,V}(y_V) + a_{i,B} \text{HP}_{i,B}(y_B), \quad (2.34)$$

où $\text{HP}_{i,j}$ correspond à un filtre passe haut appliqué sur le canal j et ajouté au canal i . Dans [Guichard *et al.*, 2009] les coefficients $a_{i,j}$ et les filtres $\text{HP}_{i,j}$ sont déterminés par calibrage en fonction de paramètres tels que la position dans l'image, la profondeur ou l'illumination.

Nous avons testé la méthode du transfert de fréquence sur les images obtenues avec l'imageur chromatique. Cependant, nous fixons les poids sur chaque canal en utilisant la carte de profondeur obtenue avec l'algorithme SAC-GLDFD. De plus, le filtrage haute fréquence est identique dans les trois canaux. Ainsi, l'équation (2.34) devient :

$$y_{i,\text{out}} = y_{i,\text{in}} + a_{p,R} \text{HP}(y_R) + a_{p,V} \text{HP}(y_V) + a_{p,B} \text{HP}(y_B). \quad (2.35)$$

Les coefficients $a_{p,i}$ sont choisis de manière très simple en fonction de la valeur de la profondeur p de la carte de profondeur. Ainsi, $a_{p,i}$ est élevé lorsque p est proche du plan de mise au point du canal i , puis décroît par morceaux lorsque p s'éloigne de ce plan. La figure 2.37 illustre les valeurs des coefficients choisis.

La figure 2.38 présente les résultats obtenus en utilisant les cartes de profondeur présentées dans le chapitre 2 page 89. Il existe une nette amélioration de l'image RVB après un transfert de fréquence, même avec un choix très basique des coefficients $a_{d,B}$ sans calibrage. Une étude plus approfondie du choix de ces coefficients permettrait certainement d'améliorer encore la résolution de l'image restaurée. Nous pouvons toutefois remarquer que la restauration de l'image présente des discontinuités locales qui sont liées aux variations locale de la carte de profondeur. Une régularisation de cette carte permettrait de réduire ces discontinuités.

Le transfert de fréquence est donc une piste intéressante pour améliorer la qualité de l'image RVB. Cette méthode est très rapide et nous avons obtenu un gain apparent de la qualité de l'image, même avec une implémentation très basique de cette méthode. Un

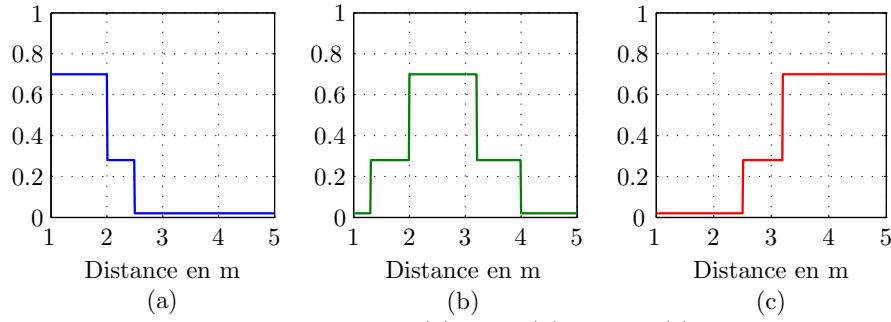


Fig. 2.37: Variation des coefficients (a) $a_{p,B}$, (b) $a_{p,V}$ et (c) $a_{p,V}$ avec la profondeur.

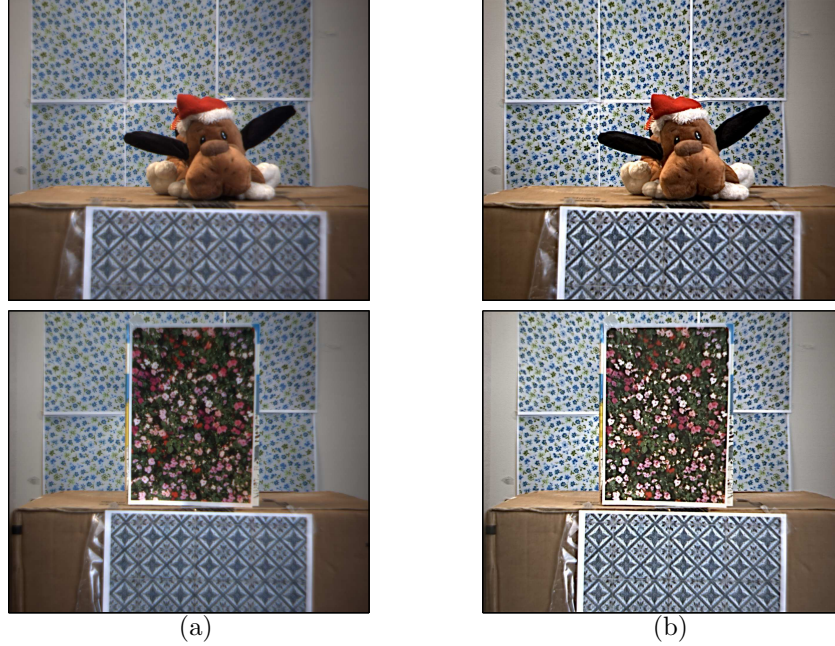


Fig. 2.38: (a) Image RVB non corrigée. (b) Image RVB après transfert de fréquence suivant l'équation (2.35).

choix plus fin des coefficients de pondérations lors du transfert de fréquence, ainsi qu'une régularisation de la carte de profondeur sont deux aspects à développer pour améliorer encore la qualité de l'image couleur.

Il faut toutefois noter qu'il existe des profondeurs pour lesquelles les trois canaux sont flous. C'est le cas notamment pour des profondeurs très proches ou très éloignées de l'imageur. Dans ce cas, le transfert de fréquence ne sera pas performant et il faut alors envisager un traitement de déconvolution. Grâce à l'estimation préalable de la profondeur et donc de la FEP, cette déconvolution n'est pas aveugle, ce qui facilite grandement la restauration. Une méthode de restauration d'image couleur issue d'un détecteur couleur présentant un filtre de Bayer, comme celle proposée dans [Soulez et Thiébaud, 2009], peut alors être envisagée.

2.5 Conclusion

Dans ce chapitre nous avons évalué expérimentalement deux prototypes d'imageur associant nos algorithmes et des optiques favorisant l'estimation de profondeur. Une première contribution est la caractérisation des performances d'estimation de profondeur d'un imageur muni d'une pupille codée sur des images réelles et des cartes de profondeurs comparées à une vérité terrain. Avec notre prototype, la pupille codée proposée dans [Levin *et al.*, 2007b] améliore la précision d'estimation de profondeur par rapport à un imageur classique, et l'écart-type expérimental varie entre 5 cm à 1.8 m et 10 cm à 2.8 m.

Mais la contribution principale de ce chapitre est le développement d'un imageur 3D exploitant le chromatisme. Cette solution avait été proposée dans la littérature, mais n'avait à notre connaissance abouti à aucune réalisation expérimentale démontrée sur des images réelles couleurs. Pour ce faire, nous avons étendu l'algorithme GLDFD tout d'abord dans le cas d'acquisitions multiples ce qui a donné lieu à l'algorithme MA-GLDFD, puis dans le cas du traitement d'images produites par un imageur chromatique, ce qui a donné lieu à l'algorithme SAC-GLDFD. La différence principale entre ces deux extensions concerne la prise en compte de la corrélation partielle des scènes à l'origine des images RVB dans le cas SAC-DFD. L'algorithme SAC-GLDFD a été validé sur des images simulées puis sur des images réelles produites par un prototype d'imageur chromatique développé en collaboration avec l'IOGS. Ces tests ont permis d'obtenir un écart-type d'estimation de profondeur sur l'axe inférieur à 5 cm autour de 1 m et de 10 cm à 3 m. L'estimation de profondeur fonctionne aussi bien pour des scènes couleurs que pour des scènes en noir et blanc, et pour différents types de texture, ce qui illustre la robustesse de l'algorithme SAC-GLDFD.

L'objectif de ce chapitre n'était pas de comparer directement les précisions obtenues avec les deux solutions optiques, car les caractéristiques des deux prototypes utilisés sont très différentes en termes de distance focale, d'ouverture et de qualité image et il est délicat de faire une comparaison qui puisse être juste. L'objectif principal de ce chapitre était de valider des concepts optiques proposés dans la littérature. Cependant, nous pouvons toutefois remarquer que l'imageur chromatique a permis d'obtenir une précision acceptable sur une plus large plage de profondeur et pour une variété plus importante de scènes que l'imageur à pupille codée. De plus, l'imageur chromatique est moins encombrant que l'imageur à pupille codée, composé d'un appareil photographique numérique.

Dans ce chapitre a été réalisée une évaluation expérimentale des performances d'estimation de profondeur de deux prototypes de nature différente. Cependant, pour concevoir un imageur à capacité 3D, des comparaisons empiriques entre de nombreuses configurations expérimentales sont difficilement envisageables. Un outil théorique de modélisation de la précision d'un imageur est nettement préférable, si il permet de prédire à moindre coût et avec une bonne fiabilité la précision d'un imageur monovoie passif achromatique ou chromatique, pour différentes configurations. Le travail de modélisation de la performance d'estimation de profondeur fait l'objet du chapitre 3.

Modélisation de la précision d'estimation de profondeur d'un imageur monovoie passif

Dans le chapitre 2, nous avons étudié les performances d'estimation de profondeur empiriques de deux prototypes, à savoir un imageur classique muni d'une ouverture codée et un imageur chromatique. Les mesures de la précision empirique ont été effectuées dans une configuration donnée de chaque imageur, or pour faire de la co-conception nous avons besoin de pouvoir comparer les performances d'imageurs en faisant varier librement des paramètres tels que la mise au point, le nombre d'ouverture ou la forme de la pupille. Une étude empirique de chaque configuration n'est donc pas envisageable. C'est pourquoi nous avons développé un modèle de performance qui permet de caractériser la précision d'estimation de profondeur théorique d'une large famille d'imageurs. Ce modèle est présenté dans ce chapitre. Il repose sur un calcul original de la borne de Cramér Rao (BCR) associée à l'estimation de la profondeur en DFD dans le cadre défini au chapitre 1. Ce modèle est présenté pour le cas d'une seule acquisition (SA-DFD), de multiples acquisitions de la même scène (MA-DFD) ou d'une acquisition produite par un imageur chromatique (SAC-DFD). Tout au long du chapitre nous présentons des expérimentations à l'appui des résultats théoriques, afin de vérifier le caractère prédictif de notre modèle au-delà du modèle a priori gaussien, sur des scènes naturelles. Nous montrons alors que si la borne théorique n'est pas atteinte, les configurations les plus favorables selon la théorie correspondent effectivement à celles qui obtiennent la meilleure précision expérimentale. Ces résultats nous incitent à utiliser notre modèle de performance dans une phase de conception d'un imageur pour des comparaisons entre plusieurs configurations ou réglages. Nous présentons deux comparaisons de ce type en fin de chapitre, illustrant le potentiel de la SAC-DFD.

ORGANISATION DU CHAPITRE

Dans la première partie de ce chapitre, nous présentons les approches développées dans la littérature pour modéliser la performance des systèmes estimant la profondeur à partir du flou de défocalisation. Ces approches ont pour défauts qu'elles ne permettent pas de modéliser des cas tels que l'utilisation d'une pupille codée ou le cas d'un imageur chromatique et qu'elles ne prennent pas en compte les aspects algorithmiques de l'estimation. Nous proposons ensuite plusieurs méthodes de calcul de la BCR dans le cas SA-DFD et puis nous utilisons cette borne pour interpréter la précision d'estimation de profondeur d'un

imageur conventionnel. Les prévisions de performance théoriques sont ensuite comparées à l'écart-type expérimental obtenu avec l'algorithme GLDFD, décrit dans le chapitre 1, sur des images simulées puis sur des images réelles produites par un imageur conventionnel. Nous étudions également, d'un point de vue théorique et expérimental, la précision d'estimation de profondeur obtenue avec un imageur muni d'une pupille codée. Nous étendons ensuite la modélisation de la précision développée dans le cas SA-DFD aux cas MA-DFD puis SAC-DFD, en comparant, dans la mesure du possible, les valeurs de l'écart-type minimal théorique donné par la BCR avec celles de l'écart-type empirique obtenues par les algorithmes MA-GLDFD et SAC-GLDFD sur des images simulées et réelles. Certains compléments permettant d'affiner le calcul de la BCR sont présentés ensuite, et nous montrons que leur effet peut être négligé dans une première analyse. La dernière partie de ce chapitre aborde la comparaison théorique de la performance d'estimation de profondeur de plusieurs imageurs qui souligne l'intérêt de l'estimation par SAC-DFD.

3.1 État de l'art

Dans cet état de l'art sont cités les articles présentant des modèles de la précision d'estimation de profondeur des méthodes reposant sur l'estimation du flou de défocalisation. Les modèles informationnels tels que ceux utilisés dans [Martinello et Favaro, 2011b; Levin *et al.*, 2007b; Zhou et Nayar, 2009a] ne sont pas considérés ici car si ils permettent de comparer la performance de différents systèmes, il ne s'agit pas de critères objectifs caractérisant la précision d'estimation de profondeur.

3.1.1 Variation de la FEP avec la profondeur

L'estimation de profondeur par DFD repose sur l'analyse locale du flou. Une première approche intuitive pour caractériser la précision de cette estimation consiste donc à s'intéresser aux paramètres qui gouvernent la sensibilité de la FEP à la profondeur. Rappelons que la taille géométrique de la FEP s'écrit :

$$\epsilon = 2Rd_{det} \left(\frac{1}{f} - \frac{1}{p} - \frac{1}{d_{det}} \right). \quad (3.1)$$

La dérivée de ϵ par rapport à la profondeur, ramenée à une grandeur en pixel, s'écrit :

$$\frac{\partial \epsilon / p_x}{\partial p} = \frac{2Rd_{det}}{p_x p^2} \quad (3.2)$$

$$= \frac{f^2}{F_{/\#} p_x p^2} \frac{p_0}{p_0 - f} \quad (3.3)$$

$$\simeq \frac{f^2}{F_{/\#} p_x p^2}, \quad (3.4)$$

en considérant $f \ll p_0$ et en notant $F_{/\#}$ le nombre d'ouverture de l'imageur, défini comme le rapport entre la distance focale et le diamètre de la lentille, dans une représentation simplifiée de l'imageur. Nous pouvons montrer que la variation de la FEP avec la profondeur est directement liée à une notion très utilisée en photographie qui est la profondeur de champ. Cette grandeur est définie comme la distance par rapport au plan de mise au point à partir de laquelle il est possible de distinguer l'apparition du flou de défocalisation, autrement dit il s'agit de la région de l'espace autour du plan de mise au point, dans laquelle un objet sera vu net par l'imageur. Une règle d'ingénieur consiste à définir la profondeur

de champ comme la distance à partir de laquelle le diamètre de la FEP dépasse un pixel. Or il existe deux distances, p^+ et p^- , situées de part et d'autre du plan de mise au point pour lesquelles $p_x < |\epsilon|$, où p_x est la taille du pixel et ϵ est donné par la formule (3.1). En utilisant les formules d'optique géométrique et en supposant que la distance focale est négligeable devant la mise au point, soit $f \ll p_0$, la profondeur de champ s'écrit :

$$PdC = p^+ - p^- \simeq \frac{2F_{/\#} p_x p^2}{f^2}. \quad (3.5)$$

Ainsi, la sensibilité de la taille géométrique de la FEP avec la profondeur peut s'écrire en fonction de la profondeur de champ :

$$\frac{\partial \epsilon / p_x}{\partial p} \simeq \frac{2}{PdC}. \quad (3.6)$$

L'équation (3.6) implique donc que pour maximiser la variation de la FEP avec la profondeur, il faut diminuer la profondeur de champ. Ainsi, si nous considérons que la précision est uniquement liée à la variation de la FEP avec la profondeur, pour augmenter la précision, il faut donc diminuer la profondeur de champ.

Dans [Blayvas *et al.*, 2007] la précision d'estimation de profondeur est associée à la distance minimale à partir de laquelle il est possible de distinguer l'apparition du flou de défocalisation. Ceci revient à associer la précision d'estimation du DFD à celle du DFF. Cette distance minimale est définie en modélisant le flou de l'image comme la convolution du flou de défocalisation, du flou lié à l'échantillonnage du détecteur et du flou lié à la diffraction. Le flou de défocalisation est modélisé simplement par un disque dont la largeur correspond à la taille géométrique ϵ de l'équation (3.1). Selon Blayvas, il est possible de distinguer un flou entre une image nette et une image défocalisée lorsque la fréquence caractéristique du flou de défocalisation dépasse celle de l'échantillonnage du détecteur et de la diffraction, cependant le choix des fréquences caractéristiques n'est pas justifié. Dans [Blendowske, 2007], les auteurs reprennent les formules développées dans [Blayvas *et al.*, 2007] en définissant les fréquences caractéristiques à l'aide des fréquences de coupure classiques, ce qui donne :

$$\Delta_{p,min} = 1.22 \frac{p^2 F_{/\#}}{f^2} \sqrt{p_x^2 + (\lambda F_{/\#})^2}. \quad (3.7)$$

Où $\Delta_{p,min}$ est la distance minimale détectable. Si les modélisations présentées dans [Blayvas *et al.*, 2007; Blendowske, 2007] sont plus précises que l'approche purement géométrique aboutissant à l'équation (3.6) et tiennent compte des effets de la diffraction, nous retrouvons la même dépendance de la précision envers les paramètres optiques de l'imageur.

Selon nous, l'approche consistant à caractériser la précision d'estimation de profondeur uniquement en étudiant la variation de la FEP avec la profondeur n'est pas suffisante. En effet, nous pouvons remarquer que cette variation tend vers l'infini lorsque p tend vers 0, ce qui implique que la précision sera maximale pour les faibles distances. Or comme le montre l'équation (3.1), lorsque p tend vers 0, la taille du flou tend également vers l'infini. Il ne semble donc pas réaliste de considérer que la meilleure précision d'estimation de profondeur sera obtenue pour des faibles distances, car l'image devient alors totalement homogène ce qui empêche l'estimation de la profondeur. En outre, comme nous avons pu le constater expérimentalement dans le chapitre 2, la précision peut être améliorée en utilisant une forme

de pupille particulière ou une optique chromatique, et les modèles basés uniquement sur la variation de la FEP avec la profondeur ne peuvent en tenir compte. De plus, la position du plan de mise au point n'apparaît pas dans la précision, ce qui peut paraître surprenant car ce paramètre influence beaucoup la quantité de flou. Enfin, l'étude de la variation de la FEP avec la profondeur pour caractériser la précision en DFD ne tient pas compte de l'influence du traitement ni du bruit.

3.1.2 Influence des traitements

Dans les références [Schechner et Kiryati, 1999; Schechner et Kiryati, 2000] la précision d'estimation de profondeur est calculée en tenant compte du traitement appliqué aux images. Il s'agit d'un contexte multi-images dans lequel la profondeur est estimée par une comparaison du rapport des spectres des images à une fréquence donnée. Pour estimer la précision, les auteurs proposent d'étudier l'effet d'une perturbation du spectre de l'une des deux images. Schechner et Kiryati obtiennent alors une relation entre la fréquence utilisée pour estimer la profondeur et la variation de flou entre les deux images, rendant l'estimation la plus robuste à la perturbation. Si dans ces deux références la précision d'estimation de profondeur est définie en tenant compte du traitement appliqué aux images, le cas traité se limite au cas MA-DFD, et de plus la caractérisation de la précision est spécifique à une méthode d'estimation de profondeur dans l'espace de Fourier.

Pour modéliser la précision d'estimation de profondeur, nous avons besoin de construire un outil prenant en compte les paramètres de l'optique, et donc la variation de la FEP avec la profondeur, mais également les paramètres de traitements, tels que le nombre de données et le niveau de bruit. Cet outil doit être suffisamment générique pour ne pas dépendre d'une méthode d'estimation de profondeur et permettre de prendre en compte l'influence de la forme de la pupille. Enfin il doit permettre de modéliser les différentes techniques de DFD que nous avons considérées, à savoir les cas SA-DFD, MA-DFD et SAC-DFD. Nous proposons pour cela d'utiliser la borne de Cramér Rao.

3.1.3 Outil statistique : la borne de Cramér Rao

La borne de Cramér Rao (BCR) est un outil statistique permettant de prévoir la variance minimale d'un estimateur non biaisé [Kay, 1993]. Elle s'exprime en fonction de la vraisemblance du paramètre θ et d'un échantillon de données \mathbf{y} de la manière suivante :

$$\text{var}(\hat{\theta}) \geq \sigma_{\text{BCR}}^2 = \text{IF}(\theta)^{-1} = -E_{\theta} \left[\frac{\partial^2 \ln L(\mathbf{y}|\theta)}{\partial^2 \theta} \right]^{-1}. \quad (3.8)$$

Où $L(\mathbf{y}|\theta)$ est la vraisemblance associée au paramètre θ et IF est l'information de Fisher. Lorsque l'estimation porte sur un vecteur $\boldsymbol{\theta}$ de n paramètres, l'information de Fisher est une matrice $n \times n$ définie par :

$$\text{IF}_{k,l}(\boldsymbol{\theta}) = -E \left[\frac{\partial^2}{\partial \theta_k \partial \theta_l} \ln L(\mathbf{y}|\boldsymbol{\theta}) \right]. \quad (3.9)$$

Classiquement, la BCR est employée pour définir une borne minimale de la variance de l'estimation du paramètre $\boldsymbol{\theta}$, indépendante de l'algorithme d'estimation, dans le cadre d'une certaine modélisation des données. Nous nous intéressons ici à son utilisation pour l'estimation de la profondeur.

Il existe plusieurs références utilisant la BCR pour l'estimation de la profondeur en MA-DFD [Rajagopalan et Chaudhuri, 1997a; Rajagopalan et Chaudhuri, 1998; Rajagopalan *et al.*, 2000]. Les calculs de la BCR reposent sur une expression de la covariance des données en supposant le flou gaussien et un modèle ARMA gaussien de la scène. Dans ces articles, l'objectif est de trouver le rapport optimal des flous des deux images utilisées en MA-DFD. Le cas SA-DFD est traité mais sert uniquement à illustrer le gain en précision lorsque plusieurs images sont utilisées, sans étude de l'influence des paramètres de l'imageur et des traitements sur la précision. Une validation expérimentale de la BCR est conduite et montre une cohérence entre les valeurs de BCR et la précision expérimentale. Dans [Johnson *et al.*, 2000], la BCR est utilisée pour prévoir la précision d'estimation de profondeur en SA-DFD avec un imageur muni d'un masque de phase sinusoïdale, avec un spectre plat pour modéliser la scène.

Par ailleurs, la BCR a plus récemment été proposée pour définir la précision d'estimation de la position d'une particule en holographie dans [Fournier *et al.*, 2010]. Dans cette référence, la particule est suffisamment petite pour que son image soit directement la FEP. Les auteurs obtiennent une expression analytique de la BCR en fonction des différents paramètres optiques, à l'aide d'un logiciel de calcul symbolique. Dans un autre contexte d'imagerie de particules, la BCR est utilisée en microscopie par fluorescence dans [Aguet *et al.*, 2005] afin de connaître la précision optimale sur la position d'une nanoparticule. Enfin, dans [Greengard *et al.*, 2006], la BCR est utilisée pour comparer la performance d'estimation de profondeur d'un imageur classique et d'un imageur muni d'optiques diffractives qui forment une FEP qui a la forme d'une double hélice 3D. La coupe de cette FEP perpendiculairement à l'axe optique donne deux taches qui effectuent une rotation autour de l'axe optique en fonction de la défocalisation.

Ces différentes références illustrent l'intérêt de la BCR pour estimer la précision d'estimation de profondeur. Cet outil a été utilisé dans les cas MA-DFD et SA-DFD mais sans étude de l'influence des paramètres de l'optique et des traitements sur la précision. De plus, le cas d'une pupille codée et le cas SAC-DFD n'ont pas été envisagés. Nous proposons d'utiliser la BCR afin d'étudier la précision d'estimation de profondeur dans les cas SA-DFD, MA-DFD et SAC-DFD. Cependant, une des difficultés est la modélisation de la scène, car contrairement aux applications en microscopie, dans lesquelles les objets sont de tailles réduites, le modèle de scène intervient dans le calcul de la BCR. Dans la suite, nous proposons un calcul original de la BCR en utilisant les modèles de scènes basés sur l'a priori de gradients gaussiens utilisé dans les chapitres 1 et 2. Ces modèles simples sont assez génériques, mais comme pour toute modélisation de la réalité, il est nécessaire de comparer les résultats théoriques obtenus avec des mesures expérimentales, ce que nous faisons dans la mesure du possible dans ce chapitre, en comparant les valeurs de l'écart-type théorique σ_{BCR} avec les valeurs de l'écart-type empirique.

3.2 Cas général

Dans cette section, nous présentons la formule générale de la BCR que nous avons utilisée dans les trois cas SA-DFD, MA-DFD et SAC-DFD en exprimant l'information de Fisher en fonction de la matrice de précision des données.

3.2.1 Rappel des modèles utilisés dans le chapitre 1

Dans le chapitre 1, nous avons modélisé la densité de probabilité de la scène sous la forme :

$$p(\mathbf{x}|Q_x) \propto \exp - \frac{\mathbf{x}^t Q_x \mathbf{x}}{2}, \quad (3.10)$$

où Q_x est la matrice de précision de la scène. Dans le cas où la matrice Q_x est régulière, nous avons vu dans l'équation (1.34) que la vraisemblance s'écrit :

$$L(\mathbf{y}|\boldsymbol{\theta}) = (2\pi)^{-\left(\frac{N}{2}\right)} |Q_y(\boldsymbol{\theta})|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{y}^t Q_y(\boldsymbol{\theta}) \mathbf{y} \right), \quad (3.11)$$

où Q_y est la matrice de précision des données, inverse de la matrice de covariance. Nous avons vu que si Q_x est régulière, alors Q_y l'est également. Rappelons que Q_y est définie par :

$$Q_y(\boldsymbol{\theta}) = Q_b - Q_b H (H^t Q_b H + Q_x)^{-1} H^t Q_b. \quad (3.12)$$

avec Q_b la matrice de précision du bruit et $\boldsymbol{\theta} \triangleq \{H, Q_x, Q_b\}$.

Cependant, dans le chapitre 1, nous avons défini $Q_x = D^t D / \sigma_x^2$ avec D la concaténation des matrices de convolution associées aux dérivées horizontales et verticales d'ordre 1, $[-1 \ 1]$ et $[-1 \ 1]^t$. Dans ce cas, la matrice Q_x est singulière, ce qui implique que la matrice Q_y l'est également et nous avons montré au paragraphe 1.2.4 que la vraisemblance pouvait s'écrire :

$$L(\mathbf{y}|\boldsymbol{\theta}) = (2\pi)^{-\left(\frac{N-1}{2}\right)} |Q_y(\boldsymbol{\theta})|_+^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{y}^t Q_y(\boldsymbol{\theta}) \mathbf{y} \right), \quad (3.13)$$

où $|Q_y|_+$ correspond au produit des valeurs propres non nulles de Q_y qui s'écrit alors :

$$Q_y(\boldsymbol{\theta}) = \frac{1}{\sigma_b^2} (I - H (H^t H + \alpha D^t D)^{-1} H^t), \quad (3.14)$$

où l'on a posé $Q_b = \sigma_b^{-2} I$. Le paramètre de régularisation $\alpha = \sigma_b^2 / \sigma_x^2$ est l'inverse d'un rapport signal à bruit et nous avons alors $\boldsymbol{\theta} \triangleq \{H, \sigma_b^2, \alpha\}$.

Dans la suite de ce chapitre, nous utilisons une nouvelle notation de la matrice de précision. Au lieu de la noter Q_y , elle est notée Q_p afin de faire apparaître la dépendance envers la profondeur p . De plus, pour simplifier, nous notons la matrice de convolution H_p au lieu de H . Dans un premier temps le seul paramètre considéré est la profondeur. Mais nous montrons à la section 3.6.1 que l'extension de ce travail à l'estimation de l'ensemble des paramètres dont dépend la vraisemblance, à savoir la FEP, le rapport signal à bruit et la variance du bruit, modifie peu les courbes de performance obtenues pour l'estimation de la profondeur seule. Nous montrons dans la suite de cette section que la BCR peut s'exprimer simplement en fonction de la matrice Q_p . Pour cela, nous distinguons à nouveau le cas d'une matrice de précision des données régulière ou singulière.

3.2.2 Information de Fisher lorsque la matrice de précision est régulière

Soit \mathbf{y} un vecteur aléatoire de taille N de moyenne nulle, contenant les pixels d'une fenêtre de l'image et Q_p sa matrice de précision régulière. D'après l'équation (3.11) sa log-vraisemblance s'écrit :

$$\ln L(\mathbf{y}|p) = -\frac{1}{2} (N \ln(2\pi) - \ln |Q_p| + \mathbf{y}^t Q_p \mathbf{y}). \quad (3.15)$$

Pour calculer la BCR, la première étape consiste à calculer l'information de Fisher IF. Nous pouvons montrer qu'en dérivant deux fois la log-vraisemblance par rapport au paramètre p l'information de Fisher s'écrit :

$$\text{IF}(p) = \frac{1}{2} \text{tr} \left(Q_p^{-1} \frac{dQ_p}{dp} Q_p^{-1} \frac{dQ_p}{dp} \right), \quad (3.16)$$

où tr désigne la trace.

Démonstration

La relation (3.16) peut être démontrée en utilisant la propriété suivante :

$$\frac{d}{dp} \ln |Q_p| = \text{tr} \left(Q_p^{-1} \frac{dQ_p}{dp} \right). \quad (3.17)$$

En dérivant l'équation (3.15) nous obtenons, en utilisant la propriété (3.17) :

$$\frac{d^2}{dp^2} \ln L(\mathbf{y}|p) = -\frac{1}{2} \left(\text{tr} \left(Q_p^{-1} \frac{dQ_p}{dp} Q_p^{-1} \frac{dQ_p}{dp} \right) - \text{tr} \left(Q_p^{-1} \frac{d^2 Q_p}{dp^2} \right) + \mathbf{y}^t \left(\frac{d^2 Q_p}{dp^2} \right) \mathbf{y} \right). \quad (3.18)$$

La moyenne de \mathbf{y} étant nulle, alors pour une matrice A donnée \mathbf{y} vérifie :

$$E[\mathbf{y}^t A \mathbf{y}] = E[\text{tr}(\mathbf{y} \mathbf{y}^t A)] = \text{tr}(Q_p^{-1} A), \quad (3.19)$$

d'où :

$$\text{IF}(p) = -E \left[\frac{d^2}{dp^2} \ln L(p) \right] \quad (3.20)$$

$$= \frac{1}{2} \text{tr} \left(Q_p^{-1} \frac{dQ_p}{dp} Q_p^{-1} \frac{dQ_p}{dp} \right). \quad (3.21)$$

■

Notons que l'information de Fisher peut s'écrire de manière équivalente en fonction de la matrice de covariance R_p :

$$\text{IF}(p) = \frac{1}{2} \text{tr} \left(R_p^{-1} \frac{dR_p}{dp} R_p^{-1} \frac{dR_p}{dp} \right). \quad (3.22)$$

3.2.3 Information de Fisher lorsque la matrice de précision est singulière

Dans le chapitre 1 nous avons montré que dans le cas où la matrice de précision des données n'est pas inversible l'inférence sur le paramètre p pouvait se faire par une projection particulière des données vers un sous-espace défini par les vecteurs propres de l'image de Q_p . Soit \mathbf{y}_p le vecteur aléatoire défini par $\mathbf{y}_p = A\mathbf{y}$, où A est une transformation linéaire des données vers l'image de Q_p , indépendante de p . Comme démontré à la section A.2 de l'annexe A, la matrice de précision de \mathbf{y}_p est définie par : $Q_{\mathbf{y}_p} = A^{+t} Q_p A^+$, avec A^+ la matrice pseudo-inverse de A . Le vecteur \mathbf{y}_p jouant le rôle de *statistique suffisante*, l'information de Fisher peut s'écrire à partir de l'équation (3.16) de la manière suivante :

$$\text{IF}(p) = \frac{1}{2} \text{tr} \left((A^{+t} Q_p A^+)^{-1} A^{+t} \frac{dQ_p}{dp} A^+ (A^{+t} Q_p A^+)^{-1} A^{+t} \frac{dQ_p}{dp} A^+ \right). \quad (3.23)$$

Un exemple de matrice de projection dans le cas d'une seule valeur propre nulle est la matrice A suivante :

$$A = \begin{bmatrix} -1 & 1 & 0 & \dots & \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \end{bmatrix}. \quad (3.24)$$

Un autre exemple est proposé dans [Besag et Kooperberg, 1995]. Il s'agit de la matrice $A = [I | -1]$ et sa pseudo-inverse $A^+ = [I | 0]^t$. Enfin, en choisissant la matrice A comme une matrice formée des vecteurs de la base de l'image de Q_p la matrice $A^+(A^+Q_pA^+)^{-1}A^+$ correspond alors à la pseudo-inverse de Q_p [Champagnat, 2012]. Ceci permet d'obtenir une expression plus intrinsèque de l'information de Fisher :

$$\text{IF}(p) = \frac{1}{2} \text{tr} \left(Q_p^+ \frac{dQ_p}{dp} Q_p^+ \frac{dQ_p}{dp} \right). \quad (3.25)$$

L'équation (3.25) montre donc que nous pouvons calculer la BCR directement à partir de la matrice de précision des données.

3.2.4 Calcul numérique de la BCR

Le calcul de l'information de Fisher par les équations (3.23) ou (3.25) fait intervenir la dérivée de Q_p par rapport à p . Nous proposons d'effectuer cette dérivée de manière numérique, en calculant les matrices $Q_p, Q_{p-\delta}$ et $Q_{p+\delta}$, avec δ un écart faible par rapport à la profondeur p , et en approchant la dérivée de Q_p par la relation :

$$\frac{dQ_p}{dp} \simeq \frac{Q_{p+\delta} - Q_{p-\delta}}{2\delta}. \quad (3.26)$$

Les matrices $Q_p, Q_{p-\delta}$ et $Q_{p+\delta}$ sont obtenues en calculant les matrices $H_p, H_{p-\delta}$ et $H_{p+\delta}$. D'après l'équation (3.14), Q_p dépend de la FEP, de α et de σ_b^2 , cependant ce dernier paramètre disparaît de la BCR dans le produit de la pseudo-inverse de Q_p et de sa dérivée par rapport à p . La BCR peut donc être calculée pour une profondeur et pour un niveau de bruit donnés.

3.3 Étude de la performance d'estimation de profondeur dans le cas SA-DFD

Cette section est dédiée à l'étude de la précision d'estimation de profondeur à partir d'une seule acquisition. Nous présentons tout d'abord trois méthodes de calcul de la BCR puis nous étudions la variation de l'écart-type théorique σ_{BCR} dans le cas d'un imageur conventionnel, en fonction des paramètres optiques et algorithmiques, avec une interprétation physique des résultats obtenus. Le calcul de σ_{BCR} est ensuite comparé à la précision empirique obtenue par l'application de l'algorithme d'estimation de profondeur GLDFD, décrit dans le chapitre 1, sur des images simulées et réelles réalisées avec un imageur conventionnel. Enfin, nous comparons les performances empiriques et théoriques d'un imageur à pupille codée.

3.3.1 Calcul de la BCR dans le cas SA-DFD

3.3.1.1 Calcul direct

Pour un imageur donné, la matrice H_p correspond directement à la matrice de convolution construite à partir de la FEP. À partir des trois matrices $H_{p-\delta}$, H_p et $H_{p+\delta}$, le calcul direct de la BCR consiste à calculer pour chaque profondeur les trois matrices $Q_{p-\delta}$, Q_p et $Q_{p+\delta}$, ce qui implique d'après l'équation (3.14) d'inverser trois fois une matrice du type $(H_p^t H_p + \alpha D^t D)$. Il faut également calculer trois fois l'inverse de la matrice $A^{+t} Q_p A^+$ ou de manière équivalente la pseudo-inverse de Q_p . Typiquement pour obtenir une valeur de BCR, il faut donc compter plusieurs dizaines de secondes pour des fenêtres et des FEP de taille environ 20×20 pixels. Ainsi, le calcul direct de la BCR est coûteux en temps de calcul ce qui est rédhibitoire dans le cadre de la co-conception. Dans la suite, nous proposons d'autres modes de calcul plus efficaces.

3.3.1.2 Calcul de la matrice de covariance par préfiltrage des données

Le principe de cette méthode est de préfiltrer les données afin de les projeter dans un sous-espace dans lequel la matrice de covariance des données préfiltrées existe et possède une expression simple. Ceci permet alors de calculer la BCR d'après l'équation 3.22. Considérons dans un premier temps l'effet du filtrage des données par un filtre quelconque w . Ce filtrage implique les relations suivantes :

$$\tilde{y} = w * y = w * h * x + w * b \quad (3.27)$$

$$= h * w * x + w * b \quad (3.28)$$

$$= h * \tilde{x} + w * b, \quad (3.29)$$

où \tilde{x} désigne la scène filtrée. Avec un modèle de bruit blanc centré et gaussien de variance σ_b^2 , la fonction d'autocorrélation des données préfiltrées s'écrit alors :

$$C(\tilde{y}) = C(h * \tilde{x}) + C(w * b) = C(h) * C(\tilde{x}) + \sigma_b^2 C(w), \quad (3.30)$$

où la notation $C(g)$ désigne la fonction d'autocorrélation de la fonction g . Ainsi, la fonction d'autocorrélation des données préfiltrées dépend de quantités connues comme l'autocorrélation de h , de w mais aussi d'une quantité à préciser $C(\tilde{x})$. L'enjeu est donc de trouver un filtre w bien choisi permettant d'exprimer simplement $C(\tilde{x})$. Pour cela, nous faisons ici une courte digression au cours de laquelle nous étudions les propriétés de la variable aléatoire $x_j | x_{-j}$ correspondant à la variable aléatoire x_j , associée à la valeur d'un pixel, connaissant tous les pixels voisins. Nous calculons ici la valeur moyenne, la variance de $x_j | x_{-j}$ puis la covariance du vecteur aléatoire $(x_j | x_{-j})_{j \in \{1, N\}}$ en se basant sur les résultats de [Besag et Kooperberg, 1995].

Propriétés statistiques de $x_j | x_{-j}$

D'après [Besag et Kooperberg, 1995], soit $\mathbf{x} = (x_1, x_2, \dots, x_N)^t$ un vecteur aléatoire de densité :

$$p(\mathbf{x}) \propto \exp \left(\frac{-\mathbf{x}^t Q \mathbf{x}}{2} \right), \quad (3.31)$$

il est possible d'expliciter la structure conditionnelle de x_j sachant tous les pixels à l'exception du pixel j par :

$$x_j | x_{-j} \sim \mathcal{N} \left(\sum_k \beta_{jk} x_k, \kappa_j \right), \quad (3.32)$$

où la variable x_{-j} regroupe tous les pixels à l'exception du pixel x_j . Les coefficients β_{jk} et κ_j s'expriment alors en fonction des valeurs de la matrice de précision suivant les relations : $\beta_{jj} = 0$, $\beta_{jk} = -Q_{x_{jk}}/Q_{x_{jj}}$ et $\kappa_j = 1/Q_{x_{jj}}$.

Calcul de la moyenne et de la variance de $x_j|x_{-j}$

D'après l'équation (3.32), nous avons besoin de connaître les coefficients $Q_{x_{jj}}$ et $Q_{x_{jk}}$. Dans la section 1.2.4 page 26, nous avons défini la matrice Q_x par :

$$Q_x = \frac{D^t D}{\sigma_x^2}, \quad (3.33)$$

où D est la concaténation des matrices de convolution L_1 et L_2 associées aux dérivées horizontales et verticales d'ordre 1, $[-1 \ 1]$ et $[-1 \ 1]^t$. En développant $\mathbf{x}^t Q_x \mathbf{x}$, et en n'écrivant que les termes associés au pixel d'indice j et ses quatres plus proches voisins appartenant au voisinage noté V_j , nous obtenons alors :

$$\mathbf{x}^t Q_x \mathbf{x} = \frac{1}{\sigma_x^2} (||L_1 \mathbf{x}||^2 + ||L_2 \mathbf{x}||^2), \quad (3.34)$$

$$= \frac{1}{\sigma_x^2} \sum_j \sum_{k \in V_j} (x_j - x_k)^2 + C_1, \quad (3.35)$$

$$= \frac{1}{\sigma_x^2} \sum_j \left(4x_j^2 - 2 \sum_{k \in V_j} x_k x_j \right) + C_2, \quad (3.36)$$

avec C_1 et C_2 contenant les termes ne dépendant pas du pixel j . Or du fait de la symétrie de Q_x , nous avons :

$$\mathbf{x}^t Q_x \mathbf{x} = Q_{x_{jj}} x_j^2 + \sum_{j \neq k} 2Q_{x_{jk}} x_j x_k + C_3, \quad (3.37)$$

avec C_3 contenant les termes ne dépendant pas du pixel j . Par identification nous pouvons en déduire que :

$$Q_{x_{jj}} = \frac{4}{\sigma_x^2}, \quad (3.38)$$

$$Q_{x_{jk}} = Q_{x_{kj}} = \begin{cases} 0 & \text{si } k \notin V_j \\ -\frac{1}{\sigma_x^2} & \text{si } k \in V_j \end{cases} \quad (3.39)$$

D'après (3.38) les termes β_{jk} de l'équation (3.32) sont donc égaux à $1/4$ pour les quatre voisins de x_j à l'ordre 1 et à 0 sinon. Ainsi, $E[x_j|x_{-j}]$ correspond à la moyenne des quatre plus proches voisins de x_j . De plus, d'après (3.39) la variance de $x_j|x_{-j}$ notée κ_j vaut $\sigma_x^2/4$.

Calcul de la matrice de covariance associée aux variables $x_j|x_{-j}$

Nous pouvons montrer que pour tous les $k \neq j$:

$$E[(x_j - E[x_j|x_{-j}])(x_k - E[x_k|x_{-k}])] = \begin{cases} 0 & \text{si } k \notin V_j \\ -\frac{\sigma_x^2}{16} & \text{si } k \in V_j \end{cases} \quad (3.40)$$

où V_j est le voisinage d'ordre 1 de x_j . En effet :

$$E[(x_j - E[x_j|x_{-j}])(x_k - E[x_k|x_{-k}])] = E \left[(x_j - E[x_j|x_{-j}]) \left(x_k - \frac{1}{4} \sum_{l \in V_k} x_l \right) \right] \quad (3.41)$$

Si x_k n'appartient pas au voisinage d'ordre 1 de x_j alors l'équation (3.41) est nulle car la variable $x_j|x_{-j}$ est décorrélée des autres pixels x_k . Par contre si k appartient au voisinage d'ordre 1 de j , il existe une valeur de l telle que $l = j$ et nous pouvons écrire :

$$E[(x_j - E[x_j|x_{-j}])(x_k - E[x_k|x_{-k}])] = -\frac{1}{4}E[(x_j - E[x_j|x_{-j}])^2] \quad (3.42)$$

$$= -\frac{\sigma_x^2}{16}. \quad (3.43)$$

Choix du filtre w

Les propriétés de la variable aléatoire $x_j|x_{-j}$ nous montrent qu'avec l'a priori sur la scène que nous avons choisi, en enlevant à x_j la moyenne de ses quatre plus proches voisins, autrement dit en posant $\tilde{x}_j = x_j - E[x_j|x_{-j}]$, la matrice de covariance de \tilde{x} est très simple. En effet, cette matrice est nulle partout sauf sur la diagonale, où elle vaut $\sigma_x^2/4$ et sur les termes correspondant au voisinage d'ordre 1, où elle vaut $-\sigma_x^2/16$. Nous choisissons donc le filtre w suivant :

$$w = \begin{pmatrix} 0 & -1/4 & 0 \\ -1/4 & 1 & -1/4 \\ 0 & -1/4 & 0 \end{pmatrix}, \quad (3.44)$$

Avec ce choix de filtre l'autocorrélation des données filtrées s'écrit :

$$C(\tilde{x}) = \frac{\sigma_x^2}{4}w. \quad (3.45)$$

En l'occurrence l'autocorrélation de la scène filtrée est proportionnelle à w . La positivité de la transformée de Fourier de w considérée comme une suite assure alors que toute matrice de Toeplitz construite à partir de w sera définie positive. En réécrivant l'équation (3.30) nous obtenons :

$$C(\tilde{y}) = \frac{\sigma_x^2}{4}C(h) * w + \sigma_b^2 C(w) \quad (3.46)$$

$$= \sigma_x^2 \left(\frac{1}{4}C(h) * w + \alpha C(w) \right). \quad (3.47)$$

La matrice R_p correspondant à la covariance des données filtrées peut donc être calculée en transformant $C(\tilde{y})$ en une matrice de Toeplitz. Ce calcul de la BCR est moins coûteux que la méthode directe car la matrice de covariance se calcule rapidement à partir de l'autocorrélation de la FEP et de w et il ne reste alors plus qu'à inverser cette matrice pour pouvoir calculer l'information de Fisher suivant l'équation (3.22) et donc la BCR. Notons que le facteur σ_x^2 disparaît dans le produit $R_p^{-1}dR_p/dp$, donc à nouveau la BCR ne dépend que du paramètre α et de p .

3.3.1.3 Calcul en utilisant la décomposition en base de Fourier

Le moyen le plus rapide d'inverser les matrices H et D est de les diagonaliser en utilisant la décomposition dans le domaine de Fourier. C'est notamment le mode de calcul utilisé dans [Rajagopalan et Chaudhuri, 1998]. En faisant l'hypothèse que la scène est périodique, les matrices de convolution H et D deviennent les matrices carrées H^e et D^e de dimensions $N \times N$ qui peuvent être diagonalisées de la manière suivante :

$$H_p^e = W^* \tilde{H}_p W \quad (3.48)$$

$$D^e = \begin{bmatrix} W^* \tilde{L}_1 W \\ W^* \tilde{L}_2 W \end{bmatrix}, \quad (3.49)$$

où W est la matrice unitaire associée à la matrice de Fourier d'ordre N et $*$ est l'opérateur adjoint (transposée conjuguée) [Idier, 2001]. Dans les équations (3.48) et (3.49), \tilde{H}_p , \tilde{L}_1 et \tilde{L}_2 sont les matrices diagonales dont les éléments correspondent respectivement à la DFT de la FEP h et à la DFT des opérateurs de dérivations $[-1 \ 1]$ et $[-1 \ 1]^t$, calculés par FFT [Hunt, 1973]. Avec cette décomposition, la matrice Q_p peut alors s'écrire $Q_p = W^* \tilde{Q}_p W$ avec \tilde{Q}_p une matrice diagonale dont les coefficients sont donnés par :

$$\tilde{Q}_p(m, m) = \frac{1}{\sigma_b^2} \frac{\alpha |\tilde{D}|^2(m, m)}{|\tilde{H}_p|^2(m, m) + \alpha |\tilde{D}|^2(m, m)}. \quad (3.50)$$

où $|\tilde{D}|^2 = |\tilde{L}_1|^2 + |\tilde{L}_2|^2$. Notons qu'une des valeurs diagonales de \tilde{Q}_p est nulle, du fait de la valeur diagonale nulle de $|\tilde{D}|^2$. Dans le calcul de BCR, nous enlevons donc cette valeur nulle ce qui donne :

$$\text{IF}(p) = \frac{1}{2} \sum_{m=1, m \neq m_0}^N \left(\tilde{Q}_p^{-1}(m, m) \frac{d\tilde{Q}_p}{dp}(m, m) \right)^2. \quad (3.51)$$

où m_0 est l'indice correspondant à la valeur diagonale nulle de $|\tilde{D}|^2$. Ici la valeur propre nulle est associée à une scène uniforme, donc à la fréquence nulle, qui correspond à $m_0=1$.

De plus, la dérivée $\frac{d\tilde{Q}_p}{dp}$ peut s'exprimer en fonction de $|\tilde{D}|^2$ et $|\tilde{H}_p|^2$, soit :

$$\frac{d\tilde{Q}}{dp}(m, m) = \frac{d\tilde{Q}}{d(|\tilde{H}_p|^2)}(m, m) \frac{d(|\tilde{H}_p|^2)}{dp}(m, m) \quad (3.52)$$

$$= -\frac{1}{\sigma_b^2} \frac{\alpha |\tilde{D}|^2(m, m)}{(|\tilde{H}_p|^2(m, m) + \alpha |\tilde{D}|^2(m, m))^2} \frac{d(|\tilde{H}_p|^2)}{dp}(m, m). \quad (3.53)$$

D'où pour tout $m \neq m_0$:

$$Q_p^{-1} \frac{d\tilde{Q}_p}{dp}(m, m) = -\frac{1}{(|\tilde{H}_p|^2(m, m) + \alpha |\tilde{D}|^2(m, m))} \frac{d(|\tilde{H}_p|^2)}{dp}(m, m), \quad (3.54)$$

et finalement :

$$\text{IF}(p) = \frac{1}{2} \sum_{m=1, m \neq m_0}^N \left(\frac{1}{|\tilde{H}_p|^2(m, m) + \alpha |\tilde{D}|^2(m, m)} \frac{d(|\tilde{H}_p|^2)}{dp}(m, m) \right)^2. \quad (3.55)$$

3.3.1.4 Choix d'une méthode de calcul

Les courbes de σ_{BCR} obtenues avec les trois méthodes de calculs précédentes sont présentées à la figure 3.1, pour un imageur dont les FEP sont modélisées par des gaussiennes avec les caractéristiques données dans le tableau 3.1.

Le calcul de σ_{BCR} par la méthode directe ou par préfiltrage donnent des valeurs très proches. En effet, l'écart entre les deux courbes est inférieur à 3 %. Le calcul par transformée de Fourier a un écart plus important que celui des deux autres méthodes de calcul, de l'ordre de 10% autour du plan de mise au point. Puis lorsque la taille de la FEP augmente, la courbe de σ_{BCR} par transformée de Fourier se sépare nettement des deux autres notamment autour de 4 m et de 1 m. Ce décrochage est lié à des effets de bords, typique de la décomposition en Fourier sur des petites fenêtres, et il disparaît lorsque la taille de la fenêtre encadrant la

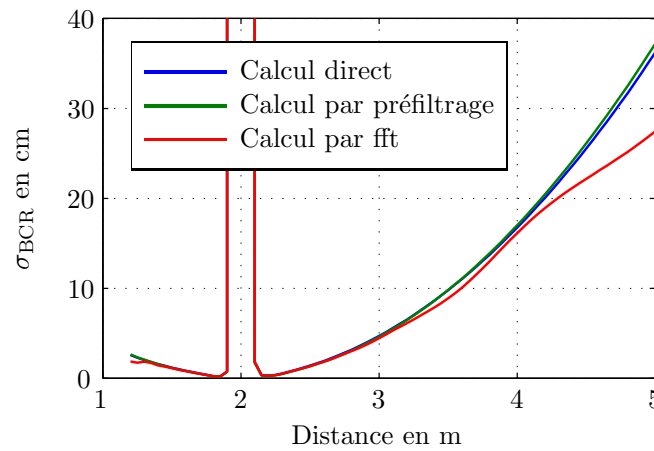


Fig. 3.1: Comparaison des valeurs de σ_{BCR} obtenues par les trois méthodes de calcul : le calcul direct, le calcul par préfiltrage des données et le calcul par la décomposition en Fourier, dans le cas de FEP gaussiennes.

FEP augmente. Ainsi, lorsque la FEP est trop large vis-à-vis de la taille de la fenêtre de FEP, des erreurs apparaissent dans la décomposition en base de Fourier. Cet effet n'est pas observé sur le calcul par préfiltrage des données, car le filtre w appliqué sur les données a un support réduit et que nous considérons des convolutions de type « valide » qui ne font aucune approximation sur les bords.

Les temps de calculs des trois méthodes sont très différents. À partir d'un jeu de trois FEP associées à une profondeur (à $p - \delta$, p et $p + \delta$) le calcul par transformée de Fourier prend 6 ms, le calcul par préfiltrage 0.13 s et le calcul direct 22 s, sous Matlab avec un processeur CPU Intel Core 2 Duo de fréquence 3GHz.

En conclusion, cet exemple montre que la méthode de calcul de la BCR la plus favorable est la méthode par préfiltrage, car elle permet un calcul relativement rapide de la BCR sur des fenêtres de petites tailles, avec un écart faible avec le calcul exact.

3.3.2 Étude de la BCR dans le cas d'un imageur conventionnel

Dans cette section, l'objectif est d'interpréter physiquement les variations de σ_{BCR} dans le cas d'un imageur conventionnel dont les FEP sont modélisées par des fonctions gaussiennes d'écart-type normalisé en pixel $\tau = \rho\epsilon/p_x$ où ϵ est défini par l'équation (3.1). Le tableau 3.1 présente les paramètres de l'imageur simulé ainsi que les paramètres algorithmiques nécessaires au calcul de la BCR.

3.3.2.1 Analyse d'un exemple de courbe de σ_{BCR}

La figure 3.2 présente la variation de l'écart-type théorique σ_{BCR} en fonction de la profondeur avec les paramètres définis dans le tableau 3.1. Notons cependant que le calcul de BCR est local, il ne permet donc pas de prendre en compte l'ambiguïté existant de part et d'autre du plan de mise au point.

L'observation des variations de σ_{BCR} avec la profondeur amène plusieurs remarques. La première est que la courbe de performance obtenue présente une divergence verticale au niveau du plan de mise au point. Cette divergence est liée à la profondeur de champ, région

f	25 mm
$F/\#$	2.5
p_x	5 μm
Mise au point	2 m
ρ	0.25
Taille des FEP	21 \times 21
α	0.001
δ	1 mm

Tableau 3.1: Paramètres utilisés pour simuler les FEP d'un imageur et calculer les valeurs de σ_{BCR} présentées à la figure 3.2.

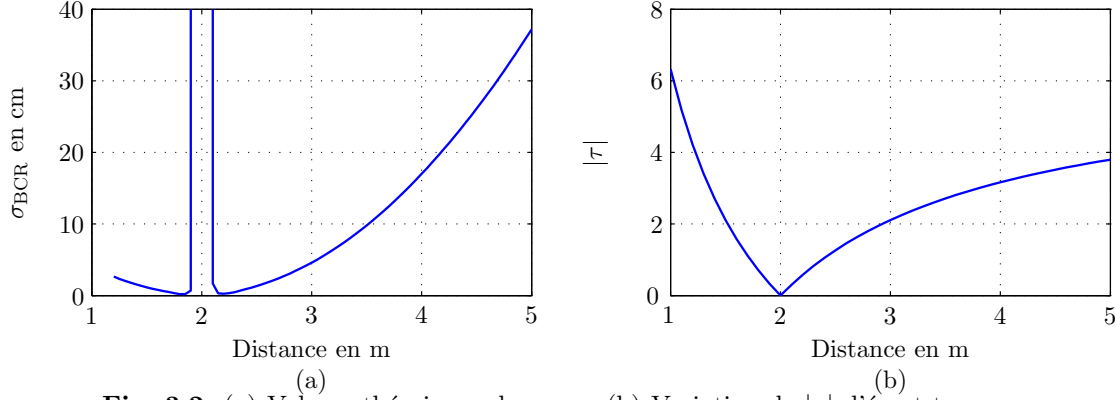


Fig. 3.2: (a) Valeurs théoriques de σ_{BCR} . (b) Variation de $|\tau|$, l'écart-type normalisé des gaussiennes modélisant les FEP, en fonction de la profondeur.

dans laquelle aucune variation de FEP n'est mesurable. En effet, la dérivée intervenant dans le calcul de la BCR est alors nulle ce qui explique la divergence de la BCR. Cet aspect du calcul numérique de la BCR est également mentionné dans [Aguet *et al.*, 2005]. Si les courbes de σ_{BCR} présentent une divergence au niveau du plan de mise au point, en pratique la précision d'estimation de profondeur n'est pas nulle dans cette région. En effet, si la FEP associée à un objet correspond à un Dirac, nous savons alors que l'objet appartient à la région définie par la profondeur de champ, même si il est impossible de mesurer une variation de la FEP. Nous pouvons alors modéliser la précision d'estimation de profondeur dans cette région en considérant que toutes les profondeurs à l'intérieur de la profondeur de champ sont équiprobables. La variance de l'estimation de profondeur correspond alors à la variance d'un tirage d'une profondeur appartenant à la profondeur de champ avec une probabilité uniforme, cette variance σ_v^2 est définie par :

$$\sigma_v^2 = \frac{PdC^2}{12} \quad (3.56)$$

L'application numérique dans l'exemple présenté à la figure 3.2 donne $\sigma_v = 46\text{mm}$. Cette modélisation approchée nous permet d'avoir un ordre de grandeur de l'incertitude sur l'estimation de profondeur au niveau du plan de mise au point.

La deuxième remarque est que d'après les variations de σ_{BCR} , la région de l'espace permettant l'estimation de profondeur la plus précise se trouve de part et d'autre du plan de mise au point, juste après la distance définie par la profondeur de champ. Ce résultat est plus réaliste que celui obtenu à la section 3.1.1, en ne considérant que la variation de

la taille de la FEP avec la profondeur, qui indiquait que la meilleure précision est obtenue lorsque la profondeur tend vers 0.

Enfin, de part et d'autre du plan de mise au point, l'écart-type semble croître de manière quadratique.

Nous avons décrit et analysé les grandes tendances de la variation du paramètre σ_{BCR} avec la profondeur sur un exemple d'imageur conventionnel simulé. Dans les paragraphes suivants, nous nous intéressons à la variation de la BCR en fonction de différents paramètres instrumentaux et algorithmiques.

3.3.2.2 Analyse de la variation de σ_{BCR} avec les paramètres algorithmiques

Les figures 3.3 (a) à (d) présentent la variation de σ_{BCR} en fonction de plusieurs paramètres algorithmiques pour l'imageur considéré précédemment, dont les FEP sont modélisées par des gaussiennes. La distance est ici fixée à 3 m et l'écart-type τ en pixel de la gaussienne modélisant la FEP vaut alors 2.1. Les figures (a) à (d) représentent respectivement la variation de σ_{BCR} en fonction de la taille du support de la FEP, du nombre de données N , du pas de dérivation numérique δ , défini dans l'équation (3.26), et de l'inverse du rapport signal à bruit caractérisé par le paramètre α .

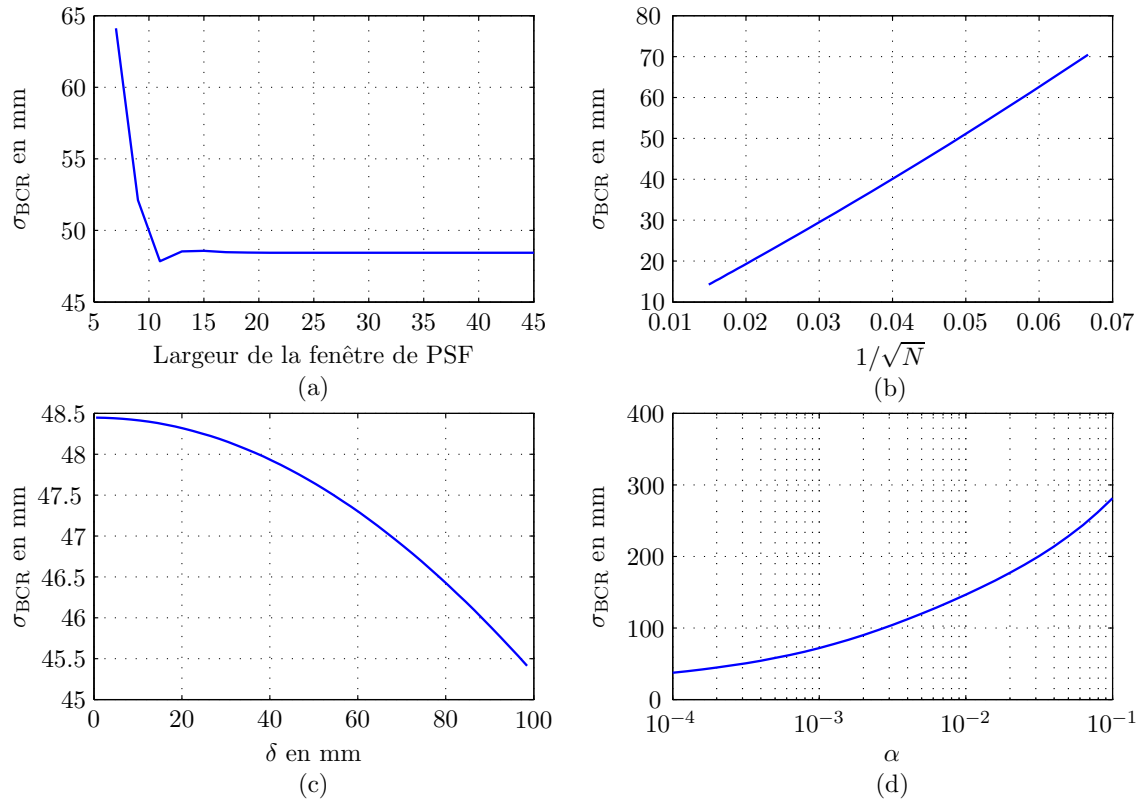


Fig. 3.3: Variation de σ_{BCR} en fonction de : (a) la largeur du support contenant la FEP, (b) le nombre de données N , (c) le pas de dérivation δ et (d) l'inverse du rapport signal à bruit α .

La figure 3.3(a) montre que la valeur de σ_{BCR} est élevée lorsque le support de la FEP, autrement dit la taille de la fenêtre contenant la FEP, est très petit, car la FEP effective est tronquée. Une fois que le support de la FEP dépasse 10 pixels, la valeur de σ_{BCR} ne varie plus. Ainsi, une fois que le support de la FEP est suffisamment large pour ne pas tronquer

la FEP, la valeur de σ_{BCR} est stable quelle que soit la taille de la fenêtre. La figure 3.3 (b) montre que σ_{BCR} est inversement proportionnel à \sqrt{N} . Rappelons que N définit la taille de l'image \mathbf{y} dans la représentation lexicographique, ce qui correspond à une fenêtre de données de taille $\sqrt{N} \times \sqrt{N}$. La figure 3.3 (c) montre que pour une valeur suffisamment faible de δ , σ_{BCR} ne varie plus sensiblement, donc le pas est suffisamment faible pour associer la dérivée analytique de la matrice Q_p à la dérivée numérique. Dans la suite, nous choisissons donc un pas de dérivation de 1 mm. Enfin, la figure 3.3 (d) montre que pour une distance fixée, σ_{BCR} augmente avec le paramètre α . Sachant qu'une augmentation de α correspond à une diminution du rapport signal à bruit, il n'est pas surprenant d'observer une diminution de la précision d'estimation de profondeur.

En conclusion, les effets des paramètres algorithmiques sur les valeurs de σ_{BCR} ne sont pas suprenants, en effet nous pouvions nous attendre à ce qu'un nombre de données et un rapport signal à bruit élevés améliorent la précision d'estimation de profondeur. De plus, la FEP ne doit pas être tronquée par le support de FEP choisi.

3.3.2.3 Interprétation de la variation de σ_{BCR} avec les paramètres optiques

Les figures 3.4(a) et (b) présentent la variation de l'écart-type théorique σ_{BCR} en fonction respectivement de la distance focale ou du rayon de la pupille R , pour une profondeur fixée et lorsque les autres paramètres optiques sont fixés. Dans le cas de la variation de la distance focale, la position d_{det} est adaptée pour conserver fixe la position du plan de mise au point p_0 . Les figures (c) et (d) présentent respectivement les variations de σ_{BCR} et de $|\tau|$ en fonction de la profondeur pour différentes valeurs de distance focale, avec un plan de mise au point fixe. De même les figures (e) et (f) présentent les valeurs de σ_{BCR} et de $|\tau|$ en fonction de la profondeur pour plusieurs valeurs de R .

Les figures 3.4(a) et (b) ont une allure similaire. Lorsque la distance focale ou le rayon de la pupille sont très faibles, l'écart-type théorique σ_{BCR} est élevé. En effet, pour ces valeurs, la FEP est inférieure à la taille du pixel, ce qui implique qu'aucune variation de la FEP avec la profondeur n'est observable. Pour des valeurs de distance focale ou de rayon plus élevées, nous observons un minimum de σ_{BCR} suivi d'une croissance d'allure linéaire de σ_{BCR} avec la profondeur. Ces deux figures montrent donc une tendance, pour une mise au point et une profondeur fixée, à favoriser soit un petit rayon de pupille, soit une faible focale, tant que la FEP n'est pas inférieure au pixel. Or d'après l'équation (3.4), la diminution de la distance focale ou du rayon de la pupille implique une diminution de la variation de la taille de la FEP avec la profondeur. Le résultat de la BCR est donc en opposition avec le raisonnement consistant à associer la précision à la variation de la taille de la FEP avec la profondeur.

Les figures 3.4(c) et (e) illustrent l'effet de la variation de la distance focale ou du rayon de la pupille sur la variation de σ_{BCR} en fonction de la profondeur. Nous observons tout d'abord qu'une distance focale plus grande ou un rayon plus grand réduit la largeur entre les deux divergences verticales. En effet, l'augmentation de ces paramètres diminue la profondeur de champ, donc la région de l'espace où la variation de la FEP n'est pas observable. Par ailleurs, pour des profondeurs éloignées du plan de mise au point, les valeurs de σ_{BCR} sont plus faibles pour une distance focale ou un rayon faible. Ce résultat est similaire à celui que nous avons observé dans les figures (a) et (b) et nous pouvons remarquer d'après les figures 3.4(d) et (f) que la configuration la plus favorable présente alors les plus petites tailles de FEP. Cependant, pour des distances proches du plan de mise au point, il existe une inversion de l'ordre des configurations les plus favorables car pour ces régions, la précision est meilleure lorsque l'imageur a un rayon ou une distance focale élevés, ce qui cette fois-ci

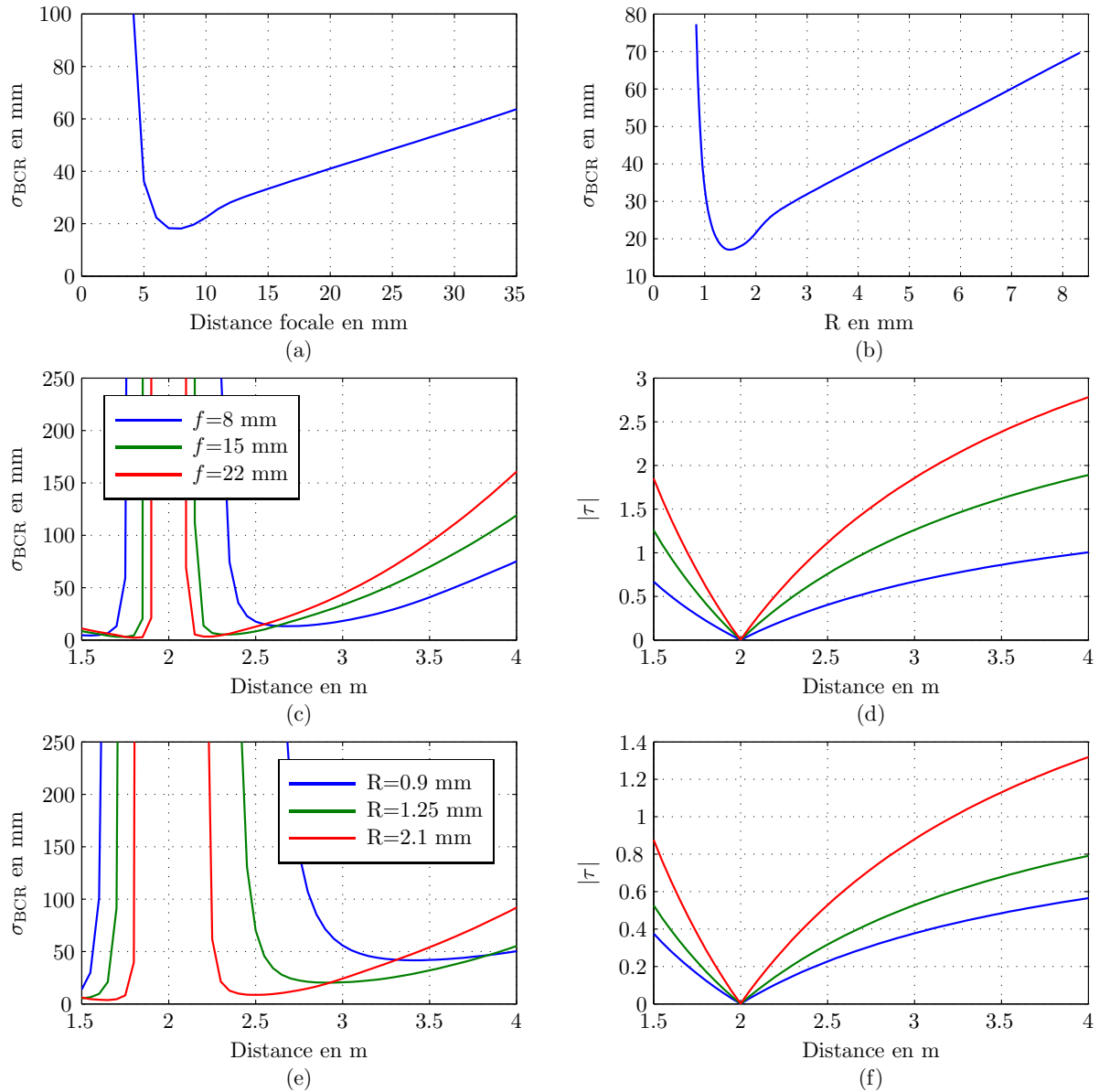


Fig. 3.4: (a) Variation de σ_{BCR} à 3 m pour un imageur dont tous les paramètres sont fixés à l'exception de la distance focale (la position du détecteur est adaptée pour avoir un plan de mise au point fixe). (b) Variation de σ_{BCR} à 3 m pour un imageur dont tous les paramètres sont fixés à l'exception du rayon de la pupille. (c) (resp.(d)) Variation de σ_{BCR} (resp. de $|\tau|$) avec la profondeur pour plusieurs valeurs de distances focales, les autres paramètres optiques étant fixés sauf d_{det} qui est adapté pour avoir une mise au point à 2 m. (e) (resp.(f)) Variation de σ_{BCR} (resp. de $|\tau|$) avec la profondeur pour plusieurs valeurs de rayon de la pupille, les autres paramètres optiques étant fixés.

est cohérent avec un critère reposant sur la variation de la taille de la FEP avec la profondeur tel que (3.4).

Finalement, la figure 3.4 illustre deux comportements contradictoires de la BCR. Loin du plan de mise au point, la meilleure précision est obtenue pour les plus faibles tailles de FEP, quitte à avoir une plus faible variation de la FEP avec la profondeur. Par contre pour des distances proches du plan de mise au point, la meilleure précision est obtenue pour une

forte variation de la FEP avec la profondeur, quitte à avoir des FEP plus larges.

Dans l'étude de la variation de la taille géométrique de la FEP avec la profondeur réalisée à la section 3.1.1, nous avons vu que la position du plan de mise au point notée p_0 n'apparaissait pas. Or, cette grandeur influence beaucoup la taille de la FEP à une profondeur donnée, nous pouvons donc penser que son influence sera significative dans les courbes de performances issues de la BCR. Les figures 3.5(a) et (b) représentent respectivement les variations de σ_{BCR} et du paramètre $|\tau|$ en fonction de la profondeur, pour un plan de mise au point situé à 2 m ou à 3 m et pour un même imageur.

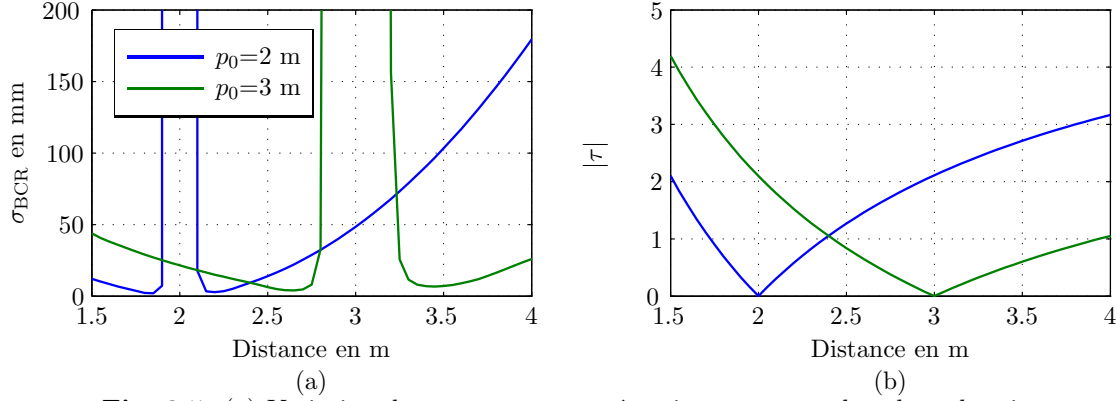


Fig. 3.5: (a) Variation de σ_{BCR} pour un même imageur avec des plans de mise au point différents. (b) Variation de τ pour les deux mises au point.

Nous observons sur la figure (a) que la valeur de σ_{BCR} dépend fortement de la position du plan de mise au point. De plus, nous observons que l'intersection des courbes de σ_{BCR} coïncident avec l'intersection des courbes de $|\tau|$, pour des profondeurs éloignées des deux plans de mise au point. Pour un imageur donné et à une profondeur donnée, la précision semble donc dépendre uniquement de la taille de la FEP.

En conclusion, d'après les courbes de performance issues du calcul de la BCR, pour des profondeurs proches du plan de mise au point, la précision semble être liée à la variabilité de la FEP avec la profondeur mais que loin des plans de mise au point elle semble liée à la taille de la FEP. Pour aller plus loin dans cette étude, nous comparons ces tendances vis-à-vis d'une formule analytique approchée de σ_{BCR} , dans le cas où les profondeurs sont éloignées du plan de mise au point. Pour cela, nous utilisons un modèle gaussien de la FEP et la décomposition en Fourier, car ils permettent une dérivation analytique simple du paramètre σ_{BCR} .

3.3.2.4 Approximation de σ_{BCR} par une formule analytique

Dans la section 3.3.1.3, nous avons vu que la matrice de précision Q_p pouvait se diagonaliser en utilisant les matrices $|\tilde{H}_p|^2$ et $|\tilde{D}|^2$ qui sont des matrices diagonales dont les coefficients non nuls correspondent respectivement au module carré de la FFT de la FEP et à la somme des modules carrés de la TFD des opérateurs de dérivation horizontale et verticale d'ordre 1, L_1 et L_2 , pris dans la représentation lexicographique. Dans cette section, nous utilisons une approximation des coefficients de ces matrices afin d'obtenir une expression analytique de la BCR.

Soient $\Delta_{\tilde{h}}$ le module carré de la TFD de la FEP et $\Delta_{\tilde{d}}$ la somme des modules carrés

de la TFD des opérateurs de dérivation L_1 et L_2 . Nous proposons alors d'utiliser les approximations suivantes :

$$\Delta_{\tilde{h}}(\mu, \nu, p) \simeq e^{-4\pi^2\tau^2(\mu^2+\nu^2)}, \quad (3.57)$$

$$\Delta_{\tilde{d}}(\mu, \nu) \simeq 4(\sin^2(\pi\mu) + \sin^2(\pi\nu)). \quad (3.58)$$

où μ et ν sont des fréquences réduites variant de $-1/2$ à $1/2-1/\sqrt{N}$. La figure 3.6 présente une comparaison entre les valeurs numériques de $\Delta_{\tilde{d}}$ et $\Delta_{\tilde{h}}$ calculées par FFT et les formules (3.57) et (3.58) pour $\sqrt{N} = 250$ et deux valeurs de τ , 3 et 0.8.

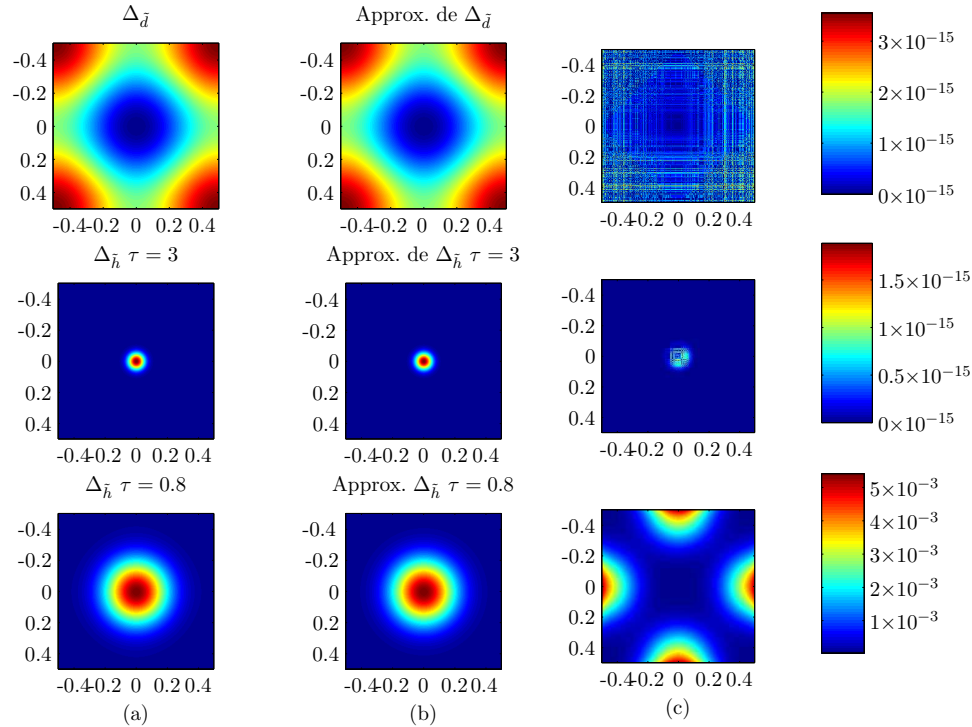


Fig. 3.6: (a) $\Delta_{\tilde{d}}$ et $\Delta_{\tilde{h}}$ calculés par FFT (b) Résultats des approximations de $\Delta_{\tilde{d}}$ et $\Delta_{\tilde{h}}$ par les équations (3.57) et (3.58). (c) Différence en valeur absolue entre l'approximation et le calcul par FFT.

L'écart entre les approximations des équations (3.57) et (3.58) et $\Delta_{\tilde{d}}$ et $\Delta_{\tilde{h}}$ calculés par FFT sont négligeables à condition que τ ne soit pas trop petit, ce qui revient à considérer des FEP larges donc des distances éloignées du plan de mise au point.

Avec les notations introduites dans cette section la BCR se calcule à partir de la relation :

$$\text{IF}(p) = \frac{1}{2} \sum_{\mu \neq 0, \nu \neq 0} \left(\frac{1}{\Delta_{\tilde{h}}(\mu, \nu, p) + \alpha \Delta_{\tilde{d}}(\mu, \nu)} \frac{d\Delta_{\tilde{h}}(\mu, \nu, p)}{dp} \right)^2. \quad (3.59)$$

Pour calculer le paramètre σ_{BCR} nous dérivons ensuite $\Delta_{\tilde{h}}$ par rapport à p ce qui donne :

$$\frac{d\Delta_{\tilde{h}}}{dp}(\mu, \nu) = \frac{d\Delta_{\tilde{h}}}{d\tau}(\mu, \nu) \frac{d\tau}{dp} \quad (3.60)$$

$$= \frac{2\rho R d_{\text{det}}}{p^2 p_x} \frac{d\Delta_{\tilde{h}}}{d\tau}(\mu, \nu) \quad (3.61)$$

$$= -\frac{2\rho R d_{\text{det}}}{p^2} 8\pi^2(\mu^2 + \nu^2) \frac{\tau}{p_x} \Delta_{\tilde{h}}(\mu, \nu). \quad (3.62)$$

En reportant l'équation (3.62) dans l'équation (3.55), σ_{BCR} peut alors s'écrire sous la forme : $\sigma_{\text{BCR}}(p) = A(p)B(p)$, avec :

$$A(p) = \frac{\sqrt{2}p^2 p_x}{16\pi^2 R d_{\text{det}}}, \quad (3.63)$$

$$B(p) = \frac{1}{\rho\tau} \frac{1}{\sqrt{\sum_{\mu,\nu} \left((\mu^2 + \nu^2) \frac{\Delta_{\tilde{h}}(\mu,\nu)}{\Delta_{\tilde{h}}(\mu,\nu) + \alpha \Delta_{\tilde{d}}(\mu,\nu)} \right)^2}}. \quad (3.64)$$

Tout d'abord le terme A peut se mettre sous la forme :

$$A(p) = \frac{\sqrt{2}p^2 F_{/\#} p_x}{8\pi^2 f^2} \frac{(p_0 - f)}{p_0}, \quad (3.65)$$

$$\simeq \frac{\sqrt{2}p^2 F_{/\#} p_x}{8\pi^2 f^2}, \quad (3.66)$$

avec $p_0 \gg f$. Nous retrouvons une expression inversement proportionnelle à la variabilité de la FEP avec la profondeur. Cela n'est pas surprenant car le calcul de la BCR fait intervenir la dérivation de la taille du flou géométrique en fonction de la profondeur. Dans l'annexe E nous montrons qu'une formule analytique approchée du terme B peut être obtenue pour des profondeurs éloignées du plan de mise au point, ce qui permet d'approcher σ_{BCR} par la formule suivante :

$$\sigma_{\text{BCR}}(p) \simeq \kappa \frac{p^2 F_{/\#} p_x}{f^2} \frac{\tau^2}{\sqrt{N} \rho (\ln \frac{\tau^2}{\alpha} - \ln \ln \frac{\tau^2}{\alpha})^{3/2}} \quad (3.67)$$

Avec $\kappa = \sqrt{6\pi}$ et

$$\tau = \frac{2R d_{\text{det}} \rho}{p_x} \left(\frac{1}{f} - \frac{1}{p} - \frac{1}{d_{\text{det}}} \right) \quad (3.68)$$

$$\tau = \frac{2f R p_0 \rho}{(p_0 - f) p_x} \left(\frac{1}{p_0} - \frac{1}{p} \right) \quad (3.69)$$

$$\simeq \frac{2f R \rho}{p_x} \left(\frac{1}{p_0} - \frac{1}{p} \right), \quad (3.70)$$

avec $f \ll p_0$. La figure 3.7 présente les courbes de σ_{BCR} obtenues avec le calcul par transformée de Fourier et la formule de l'équation (3.5) pour l'imageur décrit dans la section 3.3.2. La largeur de la fenêtre de données est fixée à $N = 10^4$.

La formule analytique (3.67) diffère dans le pire des cas de 7% par rapport aux valeurs de σ_{BCR} obtenues numériquement. Cependant elle résume la dépendance de σ_{BCR} avec différents paramètres. Tout d'abord, pour un imageur donnée, la précision d'estimation de profondeur varie en p^2 , ce qui est cohérent avec la croissance d'allure quadratique de σ_{BCR} avec la profondeur observée à la figure 3.2. La formule (3.67) fait également apparaître que σ_{BCR} s'exprime comme le produit de la précision intuitive décrite dans l'équation (3.5), liée à la variation de la FEP avec la profondeur, pondérée par un terme correctif qui tient compte de la taille de la FEP et du rapport signal à bruit.

Nous pouvons remarquer que cette formule permet de retrouver plusieurs des observations effectuées précédemment. En effet, le paramètre σ_{BCR} est inversement proportionnel à \sqrt{N} , résultat que nous avons observé numériquement à la figure 3.3 (b). De plus, la taille de la FEP, caractérisée par τ , a une influence sur les valeurs de σ_{BCR} . En considérant que le

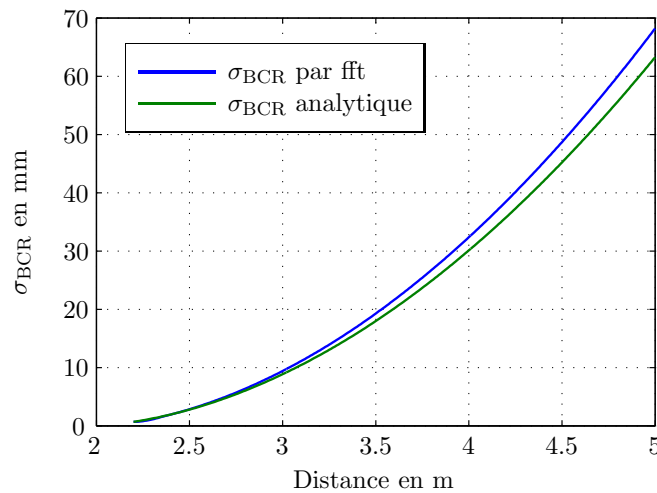


Fig. 3.7: Comparaison des valeurs de σ_{BCR} obtenues par la formule (3.67) et par le calcul par décomposition en Fourier.

terme τ^2 est prépondérant devant le terme en $(\ln \frac{\tau^2}{\alpha} - \ln \ln \frac{\tau^2}{\alpha})^{3/2}$, σ_{BCR} est proportionnelle au carré de la taille de la FEP, elle favorise donc une taille de FEP petite. Ces résultats sont cohérents avec la tendance observée aux figures 3.4 (c) à (f).

De plus, d'après l'équation (3.67), si la distance focale, la taille des pixels et le nombre d'ouverture sont fixés, alors σ_{BCR} dépend uniquement de la profondeur et de la taille de la FEP, qui dépend alors uniquement de la position du plan de mise au point. Ainsi, en considérant deux mises au point différentes d'un même imageur, si il existe une profondeur pour laquelle les tailles de FEP se croisent, alors les courbes de σ_{BCR} se croisent également. Ceci est cohérent avec l'observation faite à la figure 3.5.

De plus, en réécrivant différemment l'équation (3.67) et en utilisant à nouveau $p_0 \gg f$, σ_{BCR} peut se mettre sous la forme suivante :

$$\sigma_{BCR} \simeq \kappa \frac{p}{p_0} (p - p_0) \frac{\tau}{\sqrt{N} \rho (\ln \frac{\tau^2}{\alpha} - \ln \ln \frac{\tau^2}{\alpha})^{3/2}} \quad (3.71)$$

Notons que τ peut avoir des valeurs négatives mais le produit $(p - p_0)\tau$ est toujours positif ce qui garantit des valeurs de σ_{BCR} positives. Cette expression montre que si la profondeur et la mise au point sont fixées, la BCR dépend alors uniquement du rapport signal à bruit et de la taille de la FEP, qui peut alors être modifiée en changeant l'ouverture ou la distance focale, comme l'indique l'équation (3.70). De plus, en considérant que τ est prépondérant devant $(\ln \frac{\tau^2}{\alpha} - \ln \ln \frac{\tau^2}{\alpha})^{3/2}$, σ_{BCR} est alors proportionnel à la taille du flou de défocalisation. Ceci est cohérent avec l'allure linéaire de la variation de σ_{BCR} avec la distance focale et le rayon de la pupille R , observée numériquement aux figures 3.4 (a) et (b).

Enfin, l'équation (3.67) montre une dépendance envers le paramètre α qui n'est pas triviale, cependant elle montre que si α augmente, σ_{BCR} augmente également, ce qui est cohérent avec les observations de la figure 3.3.

3.3.2.5 Bilan

La formule analytique proposée à l'équation (3.67) est cohérente avec les observations numériques réalisées en faisant varier les paramètres de l'optique et de l'algorithme GLDFD. Elle montre que le paramètre σ_{BCR} est inversement proportionnel à la largeur de la fenêtre

de données et qu'il augmente lorsque le rapport signal à bruit diminue. De plus, σ_{BCR} est proportionnel à la variation de la FEP avec la profondeur mais dépend également de la taille du flou de défocalisation, ce qui favorise les petites tailles de FEP. Les résultats sur la dépendance de σ_{BCR} envers le nombre de données et le rapport signal à bruit ne sont finalement pas très surprenants. Cependant, le résultat sur la dépendance de σ_{BCR} avec la taille de la FEP est plus original.

Dans la suite de cette section, nous conduisons des comparaisons entre les valeurs de σ_{BCR} et l'écart-type empirique obtenu sur des images simulées et réelles à l'aide de l'algorithme GLDFD, afin de voir si le comportement théorique se retrouve expérimentalement.

3.3.3 Validations empiriques à l'aide de l'algorithme GLDFD

Après cette étude théorique de la variation de la précision en fonction de divers paramètres optiques, nous vérifions les prédictions théoriques de notre modèle à des mesures expérimentales de précision obtenues à l'aide de l'algorithme GLDFD, présenté dans le chapitre 1. Notons qu'afin de modéliser les FEP avec un modèle plus fidèle à la réalité, nous utilisons à présent les formules d'optique de Fourier présentées dans l'annexe B.

Nous avons notamment choisi de faire varier la taille de la FEP en modifiant deux paramètres qui sont la mise au point et l'ouverture, car ces deux paramètres sont directement accessibles sur un appareil photographique du commerce. Pour chaque cas, nous comparons la variation de σ_{BCR} avec l'écart-type d'estimation de profondeur empirique obtenu tout d'abord sur des images simulées et des FEP simulées en utilisant la méthode décrite dans l'annexe B, puis sur des images réelles avec des FEP calibrées. Notons que dans le cas d'images réelles, nous ne cherchons pas à comparer directement les valeurs de σ_{BCR} avec l'écart-type empirique mais simplement à vérifier les comparaisons relatives entre les différentes configurations. En effet, pour calculer la BCR nous utilisons une modélisation de la FEP pour un imageur idéal et ne tenons pas compte du sous-échantillonnage des données lié au filtre de Bayer présent sur le détecteur. De plus, l'estimation de profondeur se fait à partir de FEP calibrées, ce qui introduit une source d'erreur supplémentaire qui n'est pas modélisée dans le calcul de BCR.

3.3.3.1 Variation de la mise au point

Dans un premier temps, nous considérons un imageur ayant une ouverture et une distance focale fixée, pour deux mises au point différentes. Les FEP sont modélisées avec les caractéristiques suivantes :

f	25 mm
$F/\#$	2.5
p_x	5 μm
Taille des FEP	21 \times 21

Tableau 3.2: Paramètres utilisés pour simuler les FEP d'un imageur idéal dans le cas d'une modification de la mise au point.

Afin de modéliser une FEP polychromatique, la FEP est construite en sommant les FEP monochromatiques de trois longueurs d'onde 470 nm, 550 nm et 640 nm, choisies arbitrairement car elles correspondent aux maxima des spectres B,V et R d'un détecteur couleur.

Étude sur des données simulées

L'objectif est de comparer les valeurs de σ_{BCR} aux valeurs d'écart-type empirique mesuré à l'aide de l'algorithme GLDFD sur des données simulées. Dans un premier temps, les données simulées sont générées à partir de scènes vérifiant l'a priori gaussien, puis à partir de scènes naturelles. Pour deux positions du plan de mise au point, situé respectivement à 1.7 m ou à 2 m, une famille de FEP potentielles est simulée, correspondant à des distances comprises entre 2 et 3 m avec un pas de 10 cm. Pour les deux mises au point et pour chaque profondeur, un ensemble de 140 fenêtres d'images de taille 21×21 pixels est ensuite simulé.

Dans un premier temps, les fenêtres images sont générées en convoluant une des FEP de la famille, associée à une profondeur, avec 140 scènes générées suivant l'a priori gaussien. Un bruit blanc gaussien est ajouté aux fenêtres simulées avec une variance σ_b^2 fixée afin d'avoir $\alpha = 0.001$. L'algorithme GLDFD est alors appliqué sur chacune de ces fenêtres. La figure 3.8(a) présente l'écart-type empirique σ_e des résultats obtenus pour chaque profondeur ainsi que les valeurs de σ_{BCR} obtenues pour $\alpha = 0.001$.

Ensuite, pour chaque FEP, un ensemble de 140 fenêtres est généré en utilisant des scènes naturelles. Ces fenêtres sont simulées en convoluant 10 scènes naturelles par une des FEP de la famille et en prélevant aléatoirement 14 fenêtres sur chaque image obtenue. Les scènes naturelles utilisées sont présentées dans l'annexe F. Un bruit blanc gaussien d'écart-type 0.01 est ensuite ajouté à chaque fenêtre sachant que les scènes sont normalisées pour avoir une intensité variant entre 0 et 1. Comme nous l'avons fait dans le chapitre 2, l'écart-type des résultats d'estimation de profondeur obtenus sur ces fenêtres est évalué en enlevant 5% des valeurs extrémales, afin d'éliminer les valeurs potentiellement aberrantes. Pour comparer l'écart-type empirique obtenu avec les valeurs de σ_{BCR} , il faut connaître la valeur du paramètre α . Or contrairement au cas où les scènes sont simulées en suivant l'a priori gaussien, ici il n'existe pas de valeur théorique de α . C'est pourquoi les valeurs de σ_{BCR} sont calculées en prenant, pour chaque profondeur, la valeur moyenne des α estimés par l'algorithme GLDFD sur les 140 fenêtres. La figure 3.8(b) montre la courbe de σ_{BCR} et les valeurs d'écart-type empirique obtenues pour les deux mises au point de l'imageur.

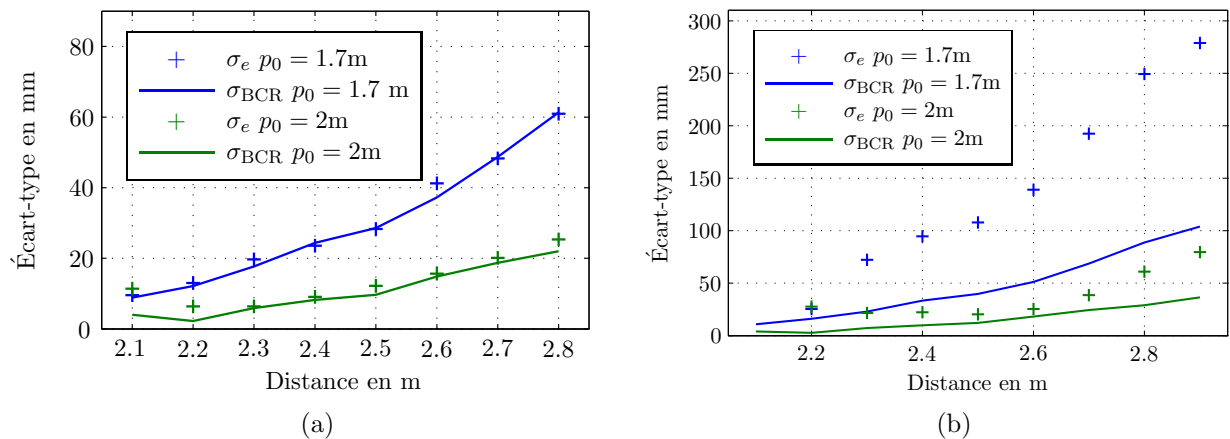


Fig. 3.8: Comparaison des valeurs de σ_{BCR} avec l'écart-type empirique σ_e des résultats d'estimation de profondeur obtenus avec l'algorithme GLDFD : (a) sur des scènes vérifiant l'a priori gaussien, (b) sur des scènes naturelles.

Pour des scènes respectant l'a priori gaussien, il y a une forte similitude entre les valeurs de σ_{BCR} et les écarts-type expérimentaux obtenus. À nouveau, il y a une différence de

précision entre les deux mises au point qui est en faveur de la mise au point à 2 m pour des distances comprises entre 2.2 et 2.8 m. Pour des scènes naturelles, l'écart-type d'estimation de profondeur empirique s'écarte plus fortement de la valeur théorique que dans la figure 3.8 (a) ce qui n'est pas surprenant car la scène réelle ne correspond pas exactement au modèle gaussien utilisé pour calculer la σ_{BCR} . Cependant la mise au point la plus favorable en termes de précision reste celle située à 2 m.

Les courbes de biais ne sont pas représentées ici mais nous vérifions, pour les deux types de scène, qu'il reste bien inférieur à l'écart-type mesuré. Nous remarquons également que le biais est plus important dans le cas de scènes naturelles que dans le cas des scènes vérifiant l'a priori gaussien utilisé pour modéliser la scène, ce qui à nouveau n'est pas surprenant.

Étude sur des données réelles

Nous utilisons le même appareil photographique Nikon que dans le chapitre 2, avec un objectif classique calibré pour deux positions du plan de mise au point différentes : à 1.5 m et à 1.8 m. La figure 3.9 présente l'écart-type empirique obtenu pour les deux mises au point pour la cible « Gauss » pour 150 fenêtres de tailles 25×25 pixels et un recouvrement de 50% de ces fenêtres. À nouveau, les valeurs de σ_{BCR} sont calculées pour chaque profondeur en prenant la valeur moyenne des α estimés par l'algorithme GLDFD sur l'ensemble des fenêtres traitées. Le tableau 3.3 présente les paramètres utilisés pour calculer les valeurs de σ_{BCR} , correspondant aux caractéristiques de l'appareil photographique.

f	35 mm
$F/\#$	2.8
p_x	6 μm
Taille des fenêtres	21×21
Taille des FEP	21×21
δ	1 mm

Tableau 3.3: Paramètres de l'appareil photographique Nikon utilisés pour calculer σ_{BCR} lors d'une modification de la mise au point.

En comparant les figures 3.9 (a) et (b) nous pouvons observer que la mise au point la plus favorable en termes d'écart-type d'estimation est la même que celle annoncée dans la théorie, à savoir la mise au point à 1.8 m. La figure 3.9 (c) montre que le biais reste inférieur à l'écart-type mesuré et qu'il est également plus faible pour la mise au point à 1.8 m.

3.3.3.2 Variation de l'ouverture

À présent nous faisons varier le nombre d'ouverture $F/\#$ d'un imageur en fixant la distance focale et la mise au point. Notons que nous nous plaçons ici dans un contexte de rapport signal à bruit fixé, afin d'étudier uniquement la variation de la taille de la pupille sur la précision. Nous étudions tout d'abord le cas d'images simulées puis le cas d'images réelles.

Étude sur des données simulées

Comme précédemment, les images simulées sont tout d'abord produites à partir de scènes vérifiant l'a priori gaussien, puis à partir d'extraits de scènes naturelles. Cependant, nous travaillons ici à une profondeur fixe en faisant varier le nombre d'ouverture entre 1.5 et 4. Pour chaque valeur du nombre d'ouverture, nous construisons une famille de FEP simulées

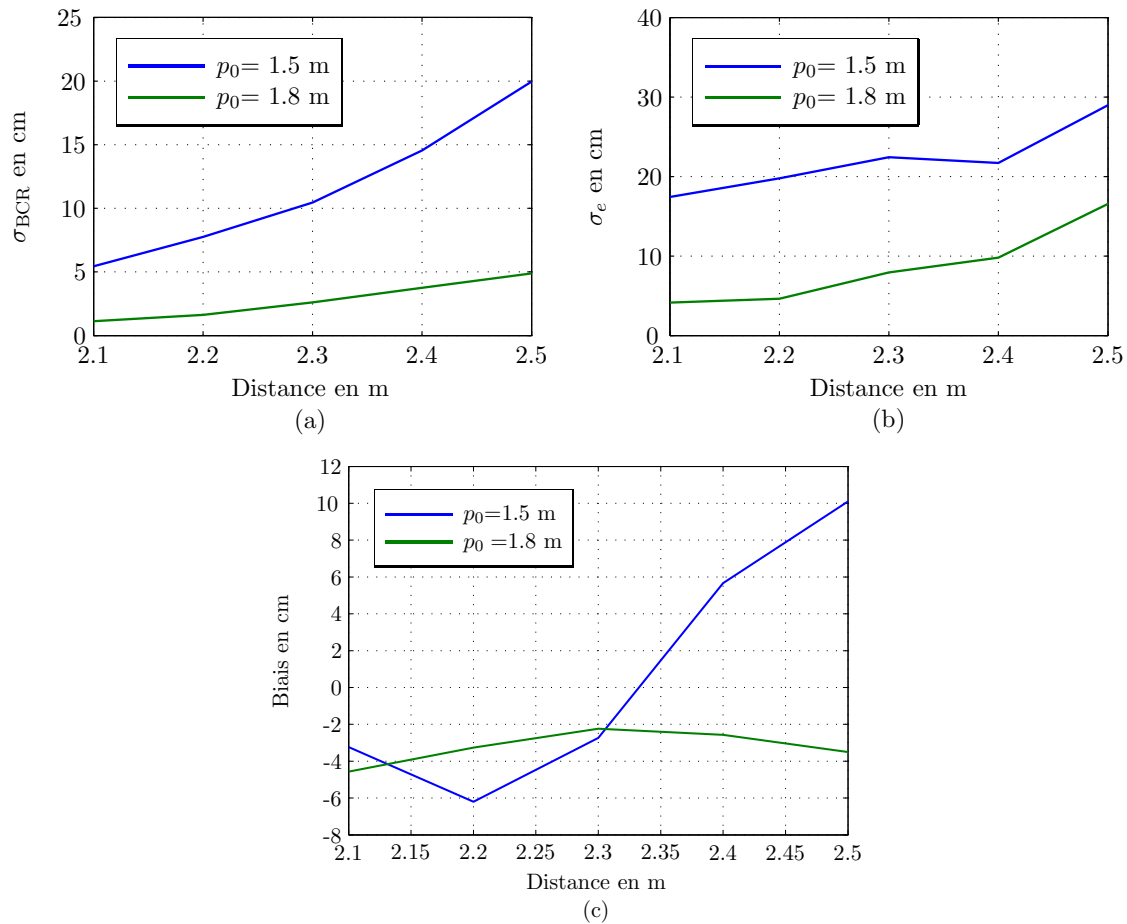


Fig. 3.9: (a) Valeurs de σ_{BCR} obtenues avec les caractéristiques du tableau 3.3 pour une mise au point à 1.5 m et une mise au point à 1.8 m. (b) Écart-type empirique σ_e obtenu sur la cible « Gauss ». (c) Biais empirique.

qui appartiennent à un intervalle de profondeur centré autour de 3 m. Les paramètres de l'imageur sont définis dans le tableau 3.1 page 108.

Pour chaque valeur d'ouverture, une série de 140 fenêtres images sont simulées en générant des scènes vérifiant l'a priori gaussien de telle sorte que $\alpha = 0.001$. L'estimation de profondeur est ensuite effectuée avec l'algorithme GLDFD sur chaque fenêtre. L'écart-type empirique σ_e obtenu pour les différentes ouvertures est présenté à la figure 3.10(a) ainsi que les valeurs de σ_{BCR} . Ensuite, pour chaque valeur d'ouverture, les fenêtres sont produites à partir de 140 extraits de scènes naturelles. Ces scènes sont présentées dans l'annexe F. Un bruit blanc gaussien d'écart-type 0.01 est ajouté à chaque fenêtre, sachant que les scènes sont normalisées. La figure 3.10(b) présente les valeurs d'écart-type empirique obtenues en enlevant 5% des valeurs extrémales. Nous représentons également les valeurs de σ_{BCR} en prenant la valeur moyenne des valeurs de α estimées par l'algorithme GLDFD sur les 140 fenêtres.

La figure 3.10 illustre que lorsque les scènes sont produites suivant l'a priori gaussien, l'écart-type empirique est proche de la valeur de précision théorique σ_{BCR} , qui favorise le nombre d'ouverture le plus élevé. Pour des scènes naturelles, l'écart-type mesuré présente un écart plus important avec les valeurs de σ_{BCR} , mais le nombre d'ouverture le plus favorable reste le nombre d'ouverture le plus grand. À nouveau les courbes de biais ne sont pas

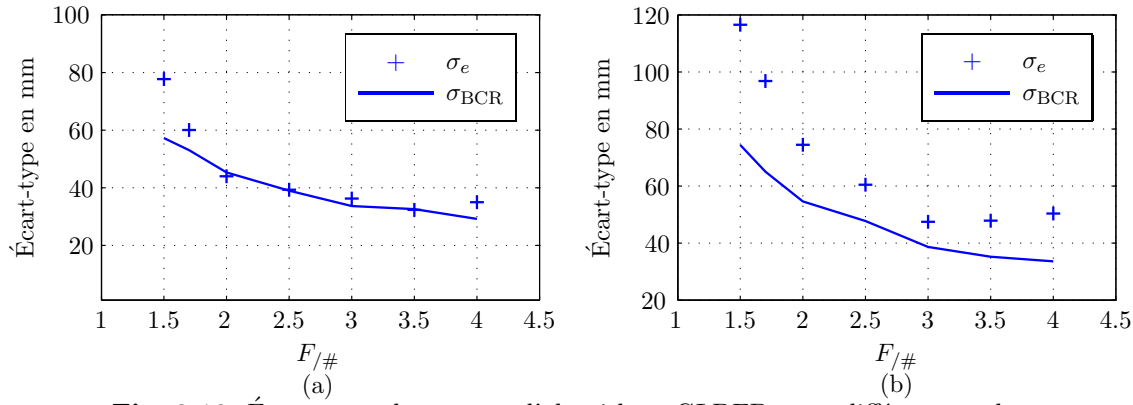


Fig. 3.10: Écart-type obtenu avec l'algorithme GLDFD pour différentes valeurs d'ouverture à 3 m. (a) Les scènes sont générées suivant l'a priori gaussien et le niveau de bruit est tel que $\alpha = 0.001$. (b) Les scènes sont générées à partir d'extraits de scènes naturelles avec un écart-type de bruit égale à 0.01. Les valeurs d'écart-type sont comparées aux valeurs de σ_{BCR} .

présentées ici, mais nous vérifions qu'il est négligeable devant les valeurs d'écart-type dans le cas de scènes vérifiant l'a priori gaussien. Ce biais augmente dans le cas de scènes naturelles mais reste bien inférieur à l'écart-type mesuré.

Étude sur des données réelles

À nouveau, nous utilisons l'appareil photographique Nikon pour effectuer une validation expérimentale des résultats précédents. Cet appareil est utilisé avec une mise au point à 1.7 m et deux ouvertures $F/\#=4$ et $F/\#=2.8$, avec un objectif muni d'une pupille classique. Le temps d'intégration est adapté pour avoir un flux identique sur chaque pixel, afin d'avoir un rapport signal à bruit identique pour les deux ouvertures. La figure 3.11(b) présente l'écart-type empirique d'estimation de profondeur obtenu pour la mire « Gaussienne » placée à différentes positions de l'appareil. La statistique des résultats est calculée au centre de l'image dans une région de 140×200 pixels avec des fenêtres de taille 25×25 pixels et un recouvrement de 50%, ce qui donne 150 mesures de profondeur. Ces résultats sont comparés aux valeurs de σ_{BCR} présentées à la figure 3.11(a), calculées pour une valeur de α correspondant à la valeur moyenne des valeurs de α estimée par l'algorithme GLDFD pour chaque profondeur et pour chaque fenêtre.

Expérimentalement, l'ouverture de 4 permet d'estimer la profondeur avec une meilleure précision que l'ouverture de 2.8, ce qui est cohérent avec la théorie. De plus, sur les courbes de σ_{BCR} , l'écart entre les deux ouvertures se réduit lorsque les profondeurs de rapprochent du plan de mise au point. Nous observons également cet effet sur l'écart-type empirique. Concernant le biais représenté à la figure 3.11(c) nous pouvons constater qu'il est très faible pour les deux ouvertures pour des profondeurs comprises entre 1.7 m et 2 m puis qu'il augmente mais reste inférieur au pas de calibrage de 5 cm. Cependant, le biais reste inférieur à l'écart-type mesuré.

Les courbes 3.11(b) indiquent, conformément à la théorie, qu'une ouverture plus petite permet une meilleure estimation de profondeur, lorsque la mise au point et la distance focale sont fixées. Notons toutefois que ce résultat considère un rapport signal à bruit fixé, or en pratique, l'utilisation d'une ouverture plus petite augmente le bruit. Ici nous avons

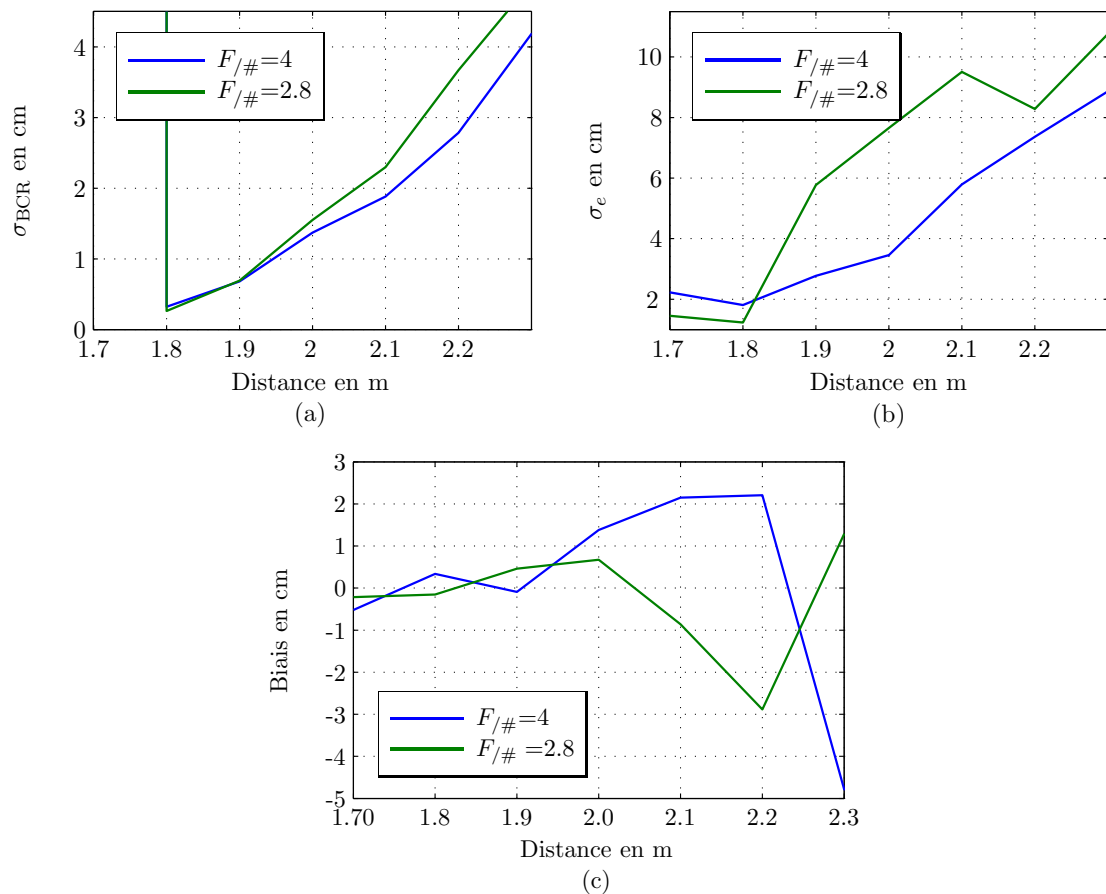


Fig. 3.11: (a) σ_{BCR} calculé à partir de FEP simulées et des caractéristiques optiques de l'objectif Nikon présenté dans le tableau 3.3. (b) Écart-type empirique σ_e pour les deux ouvertures. (c) Biais empirique.

compensé cet effet en augmentant le temps d'acquisition, afin de faire varier la BCR avec un seul paramètre qui est le rayon de la pupille. Cependant, il serait intéressant de comparer les deux pupilles avec un temps d'intégration constant afin d'évaluer le gain en précision de la pupille la plus petite, en tenant compte de la diminution du rapport signal à bruit induit par la limitation plus importante du flux de lumière.

3.3.3.3 Bilan

Dans les deux sections précédentes, nous avons comparé les valeurs de σ_{BCR} à l'écart-type empirique mesuré sur des images simulées et réelles, en utilisant l'algorithme GLDFD. Sur des images simulées, les écarts-types mesurés sont proches des valeurs théoriques mais une différence plus importante entre l'écart-type expérimental et σ_{BCR} a été observée sur des images réelles, ainsi qu'une augmentation du biais. Ceci n'est pas surprenant car d'une part les scènes naturelles ne suivent pas précisément l'a priori gaussien utilisé pour calculer σ_{BCR} . D'autre part, dans la procédure expérimentale les FEP sont calibrées ce qui introduit des erreurs supplémentaires sur l'estimation de profondeur, qui ne sont pas modélisées dans notre calcul de BCR. Cependant, dans les deux situations que nous avons étudiées, la configuration permettant l'estimation de profondeur avec la meilleure précision est identique à celle prévue par la BCR. L'étude expérimentale confirme aussi le résultat nouveau que

nous avons obtenu en étudiant la BCR théorique, à savoir que la précision d'estimation de la profondeur loin du plan de mise au point s'améliore lorsque la taille de la FEP diminue, pour un rapport signal à bruit fixé.

3.3.4 Comparaison des performances d'un imageur à pupille codée et à pupille circulaire

Dans [Levin *et al.*, 2007b] une pupille codée a été optimisée pour favoriser l'estimation de profondeur. L'objectif de cette section est tout d'abord de comparer les valeurs de σ_{BCR} avec une pupille circulaire ou avec une pupille codée, afin de voir si un gain en précision est observé également en utilisant notre modèle de performance. Puis les valeurs de σ_{BCR} sont comparées aux valeurs d'écart-type empirique obtenues sur des images simulées et réelles avec l'algorithme GLDFD.

3.3.4.1 Comportement théorique

La figure 3.12 montre les valeurs de σ_{BCR} obtenues pour un même imageur simulé, avec et sans la pupille codée d'A. Levin. Les caractéristiques de l'imageur simulé sont données dans le tableau 3.4 et nous prenons $\alpha = 0.001$.

f	25 mm
$F/\#$	2
p_x	5 μm
Mise au point	2 m
Taille des fenêtres	21 \times 21
Taille des FEP	21 \times 21
Pas dérivation	1 mm

Tableau 3.4: Paramètres utilisés pour tracer la figure 3.12 .

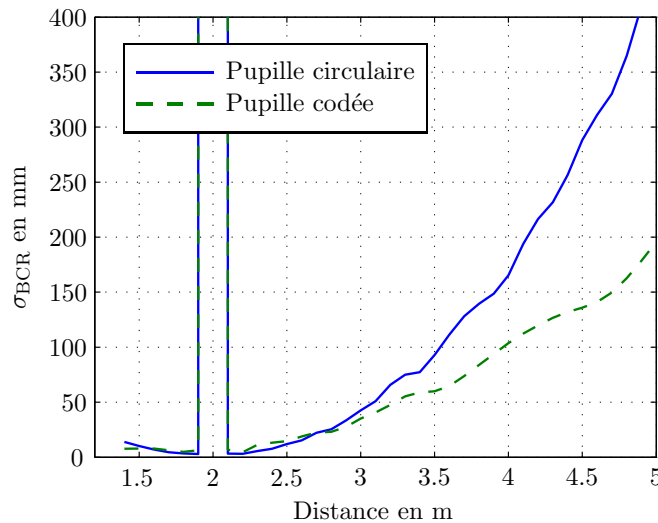


Fig. 3.12: Comparaison des valeurs de σ_{BCR} pour un imageur à pupille codée et un imageur conventionnel de mêmes caractéristiques.

La figure 3.12 montre que la pupille codée améliore les performances d'estimation de profondeur, notamment pour les grandes profondeurs. En effet, la forme de la pupille se distingue dans la FEP pour de fortes défocalisations et elle aura donc une influence notable surtout pour des profondeurs éloignées du plan de mise au point.

Notons que sur cette figure, le paramètre α est identique pour les deux pupilles. Ceci revient à ne pas tenir compte de la perte de flux liée à l'introduction de la pupille codée, c'est-à-dire en pratique à compenser la perte de flux par une augmentation du temps d'intégration.

3.3.4.2 Étude sur des données simulées

Nous comparons à présent la performance expérimentale d'estimation de profondeur avec ou sans pupille codée à l'aide de l'algorithme GLDFD, sur des images simulées. Pour cela, nous construisons deux familles de FEP potentielles à partir des paramètres définis dans le tableau 3.4, pour des profondeurs comprises entre 3 et 4 m et avec un pas de 2 cm. Dans une des familles la pupille est circulaire, dans l'autre elle a la forme de la pupille de Levin. Pour chaque type de pupille et pour chaque profondeur, 140 fenêtres images sont simulées à partir de scènes respectant l'a priori gaussien et avec l'ajout d'un bruit blanc gaussien tel que $\alpha=0.001$. L'algorithme GLDFD estime la profondeur sur chacune de ces fenêtres. L'écart-type empirique des résultats obtenus est présenté à la figure 3.13(a), ainsi que les valeurs de σ_{BCR} calculées pour une valeur de $\alpha = 0.001$.

Ensuite, pour chaque profondeur, 140 fenêtres images sont générées à partir d'extraits de scènes prélevés aléatoirement sur 10 scènes naturelles présentées dans l'annexe F, pour un écart-type de bruit fixé à 0.01, sachant que les scènes sont normalisées. La figure 3.13(b) présente l'écart-type empirique obtenu pour les deux objectifs ainsi que les courbes de σ_{BCR} calculées avec la valeur de α correspondant, pour chaque profondeur, à la moyenne des valeurs de α estimées par l'algorithme GLDFD sur les 140 fenêtres.

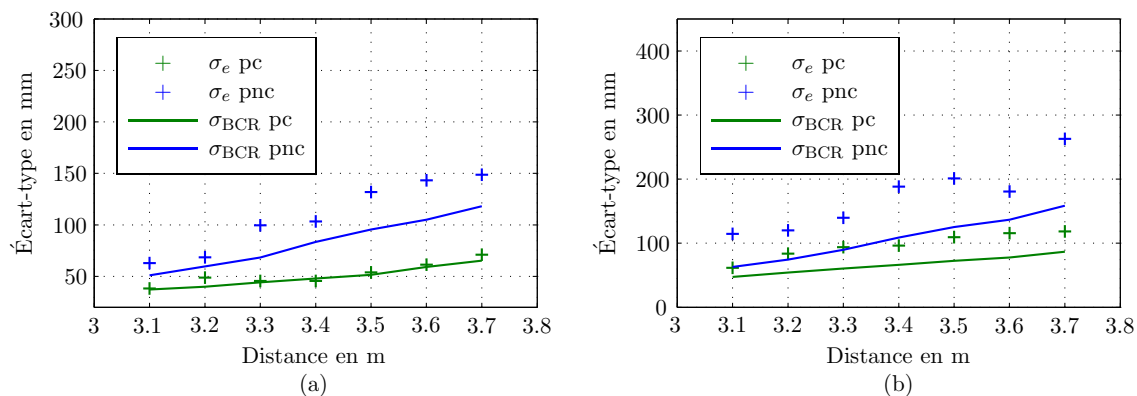


Fig. 3.13: Comparaison de σ_{BCR} et de l'écart-type obtenu avec l'algorithme GLDFD dans le cas d'un imageur à pupille codée (pc) et à pupille classique (pnc) : (a) sur des scènes respectant l'a priori gaussien et $\alpha = 0.001$, (b) sur des scènes naturelles avec $\sigma_b = 0.01$.

Lorsque les scènes sont générées suivant l'a priori gaussien les mesures d'écarts-types correspondent aux valeurs de σ_{BCR} , qui favorise l'estimation de profondeur à l'aide de la pupille codée. Nous observons alors un biais compris entre 5 mm pour les plus faibles profondeurs et 25 mm pour les profondeurs voisines de 3.6 m, ce qui reste négligeable devant l'écart-type expérimental. Dans le cas d'images produites avec des scènes naturelles, les valeurs d'écart-type empirique sont moins bonnes que ce qu'indique la BCR. Toutefois,

la pupille codée donne toujours une meilleure précision que la pupille traditionnelle. Nous pouvons observer que le biais est plus important que dans le cas précédent, mais il reste bien inférieur à l'écart-type mesuré.

3.3.4.3 Études sur des données réelles

L'appareil photographique Nikon est utilisé avec deux objectifs identiques, l'un d'eux a une pupille codée telle que celle proposée par A. Levin et l'autre a une ouverture traditionnelle. Nous reprenons ici les résultats d'estimation de profondeur obtenus dans la section 2.3.3 du chapitre 2, mais nous considérons ici les profondeurs pour lesquelles les résultats obtenus ont un biais négligeable soit pour des profondeurs comprises entre 1.8 m et 2.5 m. Nous présentons les résultats obtenus sur la cible « Gauss » présentée à la figure 2.6 de la page 62 avec les deux objectifs pour une mise au point à 1.6 m. L'ouverture de l'imageur traditionnelle est choisie pour que son diamètre corresponde au diamètre de la pupille codée, afin d'avoir des FEP de tailles identiques et le temps d'intégration est adapté afin d'avoir le même rapport signal à bruit avec les deux objectifs. La figure 3.14 présente les valeurs d'écart-types et de biais mesurées à plusieurs profondeurs avec les deux objectifs, sur 150 fenêtres appartenant à une région centrée de l'image. Nous présentons également les courbes de σ_{BCR} , calculées en prenant à chaque profondeur la moyenne des valeurs de α estimées par l'algorithme GLDFD sur l'ensemble des fenêtres traitées.

La figure 3.14(b) illustre le gain en précision mesuré expérimentalement lorsque l'objectif est muni d'une pupille codée. Ce résultat est cohérent avec la prédiction de la BCR. De même nous pouvons observer un gain significatif de la pupille codée en termes de biais, ce que nous avons pu observer dans le chapitre 2, mais qui ne peut pas être modélisé par la BCR.

Notons que comme précédemment, la comparaison directe des valeurs d'écart-type et de σ_{BCR} n'est pas possible car les valeurs de σ_{BCR} sont calculées pour des FEP simulées par les formules d'optique de Fourier sans tenir compte des aberrations ni du sous-échantillonnage des données liées au schéma Bayer du détecteur. De plus, cette comparaison est réalisée pour un rapport signal à bruit fixé, ainsi, la perte de flux lumineux liée à l'utilisation de la pupille codée a été compensée par un temps d'intégration supérieur. En pratique ce réglage peut avoir des inconvénients, par exemple un flou de bougé plus important, il serait donc à nouveau intéressant de reproduire ces tests en conservant un temps d'acquisition constant, afin de comparer les performances de la pupille codée et de la pupille classique dans les mêmes conditions d'utilisation de l'appareil photographique.

3.3.5 Bilan

L'étude précédente confirme à la fois par la BCR et par les tests expérimentaux, l'apport de l'utilisation d'une pupille codée en termes de précision d'estimation de profondeur. Cependant, cette étude ne tient pas compte de la diminution du RSB liée à la diminution du flux lumineux induit par l'utilisation d'une pupille codée.

3.4 Étude de la performance d'estimation de profondeur dans le cas MA-DFD

L'objectif de cette section est d'étendre l'étude de performance fondée sur la BCR dans le cas MA-DFD et de comparer les valeurs d'écart-type théorique σ_{BCR} à des mesures de l'écart-type empirique obtenues avec l'algorithme MA-GLDFD.

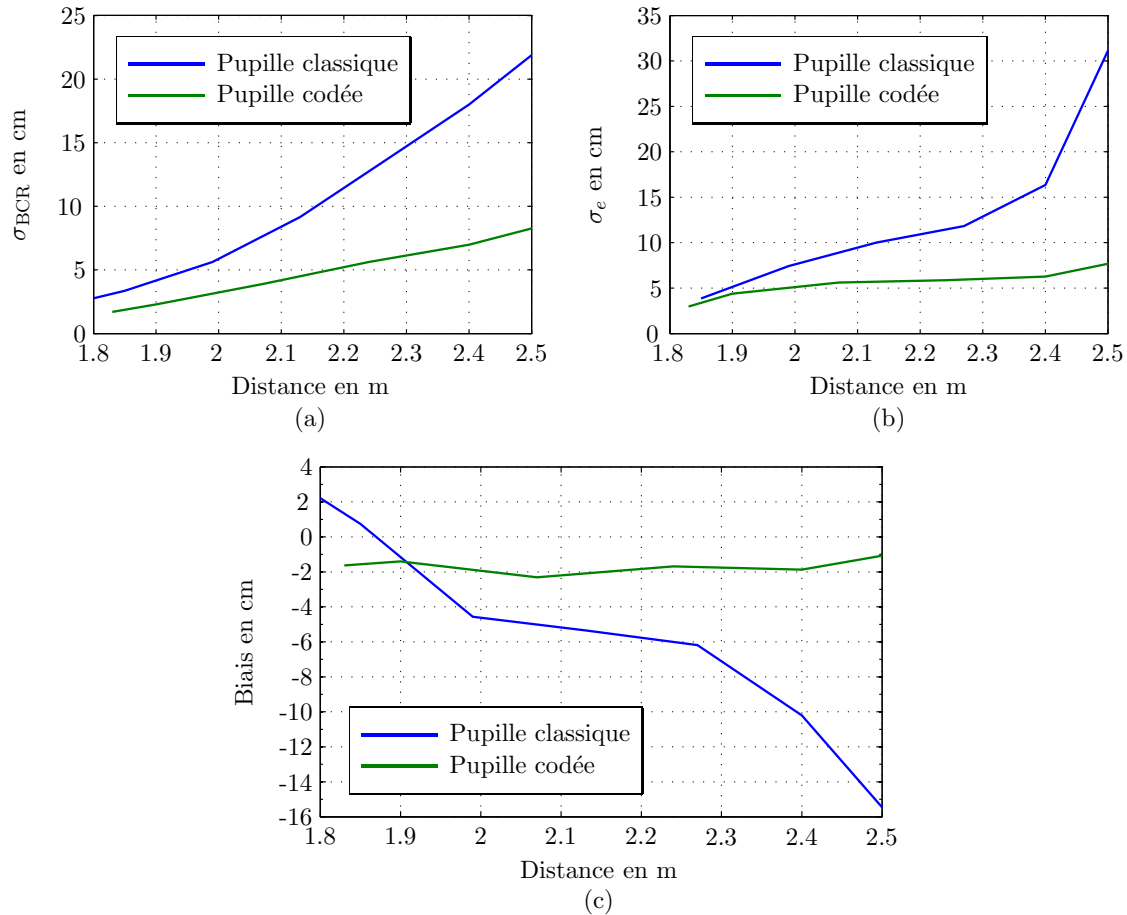


Fig. 3.14: Comparaison des performances d'un objectif avec et sans pupille codée. (a) σ_{BCR} calculé en prenant la moyenne des α estimés sur les fenêtres à une profondeur donnée. (b) Écart type empirique. (c) Biais empirique.

3.4.1 Calcul de la BCR dans le cas MA-DFD

3.4.1.1 Calcul direct

Dans le cas du MA-DFD, d'après l'expression de la vraisemblance obtenue dans le chapitre 2 à la page 75 la matrice de précision s'écrit :

$$Q_p = \frac{1}{\sigma_b^2} (I - H_M (H_M^t H_M + \alpha D^t D)^{-1} H_M^t) \quad (3.72)$$

avec $H_M = [H_{p,1}^t H_{p,2}^t \dots H_{p,K}^t]^t$ et \mathbf{Y} la concaténation des images dans la représentation lexicographique. Le calcul direct de la BCR est donc identique au cas SA-DFD en construisant préalablement la matrice H_M par concaténation des matrices de convolution associées aux FEP des différentes configurations. L'inversion de la matrice Q_p est encore plus coûteuse en temps de calcul que dans le cas SA-DFD car la matrice Q_p est de plus grande dimension.

3.4.1.2 Calcul de la matrice de covariance par préfiltrage des données

Le préfiltrage des données, proposé dans la section 3.3.1.2, peut également être utilisé dans le cas MA-DFD. Les images préfiltrées s'écrivent alors :

$$\tilde{\mathbf{y}}_1 = w * \mathbf{y}_1 = w * h_{p,1} * \mathbf{x} + w * \mathbf{b}_1 \quad (3.73)$$

$$\tilde{\mathbf{y}}_2 = w * \mathbf{y}_2 = w * h_{p,2} * \mathbf{x} + w * \mathbf{b}_2, \quad (3.74)$$

et la matrice de covariance est définie par :

$$R_p = \begin{bmatrix} R_{p,1,1} & R_{p,1,2} \\ R_{p,2,1} & R_{p,2,2} \end{bmatrix}, \quad (3.75)$$

où chaque matrice $R_{p,s,t}$ est une matrice construite à partir des matrices de corrélation :

$$C(s, s) = \sigma_x^2 \left(\frac{1}{4} C(h_s, h_s) * w + \alpha C(w) \right) \quad (3.76)$$

$$C(s, t) = \sigma_x^2 \left(\frac{1}{4} C(h_s, h_t) * w \right), \quad (3.77)$$

en considérant que les deux vecteurs aléatoires \mathbf{b}_1 et \mathbf{b}_2 sont décorrélés et en posant $C(h, g)$ l'intercorrélation des fonctions quelconques f et g . Chaque matrice de type $C(s, t)$ est ensuite transformée en une matrice de Toeplitz $R_{ps,t}$.

3.4.1.3 Calcul en utilisant la décomposition de Fourier

Comme dans le cas SA-DFD, la BCR peut être calculée en diagonalisant les matrices dans le domaine de Fourier. En faisant l'hypothèse de matrices circulantes, les deux matrices de convolution correspondant aux deux configurations de l'imageur peuvent être diagonalisées sous la forme :

$$H_M^e = \begin{bmatrix} W^* \tilde{H}_{1,p} W \\ W^* \tilde{H}_{2,p} W \end{bmatrix}, \quad (3.78)$$

où W est la matrice unitaire associée à la matrice de Fourier d'ordre N et $*$ est l'opérateur adjoint (transposée conjugué) [Idier, 2001]. Les matrices $\tilde{H}_{1,p}$ et $\tilde{H}_{2,p}$ sont les matrices diagonales dont les éléments sont la DFT des deux FEP de chaque configuration. Avec cette diagonalisation nous pouvons écrire : $Q_p = W_2^* \tilde{Q}_p W_2$ avec W_2 défini par :

$$W_2 = \begin{bmatrix} W & 0 \\ 0 & W \end{bmatrix}, \quad (3.79)$$

et

$$\tilde{Q}_p = \begin{bmatrix} \tilde{Q}_{p1,1} & \tilde{Q}_{p1,2} \\ \tilde{Q}_{p2,1} & \tilde{Q}_{p2,2} \end{bmatrix}, \quad (3.80)$$

où pour $\{s, t\} = \{1, 2\}$ ou $\{s, t\} = \{2, 1\}$:

$$\tilde{Q}_{ps,t}(m, m) = \frac{|\tilde{H}_{s,p}|^2(m, m) + |\tilde{H}_{t,p}|^2(m, m) + \alpha |\tilde{D}|^2(m, m) - \tilde{H}_{s,p} \tilde{H}_{t,p}^*(m, m)}{|\tilde{H}_{s,p}|^2(m, m) + |\tilde{H}_{t,p}|^2(m, m) + \alpha |\tilde{D}|^2(m, m)}, \quad (3.81)$$

et

$$\tilde{Q}_{ps,s}(m, m) = \frac{|\tilde{H}_{t,p}|^2(m, m) + \alpha |\tilde{D}|^2(m, m)}{|\tilde{H}_{s,p}|^2(m, m) + |\tilde{H}_{t,p}|^2(m, m) + \alpha |\tilde{D}|^2(m, m)}. \quad (3.82)$$

Rappelons que $|\tilde{D}|^2$ a été défini à la section 3.3.1.3 page 3.3.1.3. Chaque matrice $\tilde{Q}_{p,s,t}$ est une matrice diagonale, c'est pourquoi afin d'inverser \tilde{Q}_p rapidement, nous réordonnons les données afin d'avoir :

$$\tilde{Q}_p = \begin{bmatrix} \tilde{Q}_p^r(1) & 0 & 0 & 0 & 0 \\ 0 & \tilde{Q}_p^r(2) & 0 & 0 & 0 \\ 0 & 0 & \dots & \tilde{Q}_p^r(m) & 0 \\ 0 & 0 & 0 & 0 & \tilde{Q}_p^r(N) \end{bmatrix}, \quad (3.83)$$

où :

$$\tilde{Q}_p^r(m) = \begin{bmatrix} \tilde{Q}_{p1,1}(m,m) & \tilde{Q}_{p1,2}(m,m) \\ \tilde{Q}_{p2,1}(m,m) & \tilde{Q}_{p2,2}(m,m) \end{bmatrix}. \quad (3.84)$$

En effet, la BCR est invariante par un changement de base. \tilde{Q}_p est alors diagonale par bloc et pour l'inverser il suffit d'inverser chaque bloc \tilde{Q}_p^r . Finalement l'information de Fisher est calculée avec la relation :

$$\text{IF}(p) = \frac{1}{2} \sum_{m=1, m \neq m_0}^N \text{tr} \left(\tilde{Q}_p^{r-1} \frac{d\tilde{Q}_p^r}{dp} \tilde{Q}_p^{r-1} \frac{d\tilde{Q}_p^r}{dp} \right). \quad (3.85)$$

3.4.1.4 Exemple

Les trois méthodes décrites précédemment sont comparées sur un exemple simple d'imageur dont les FEP sont modélisées par une gaussienne. Les caractéristiques de l'imageur utilisées sont identiques à celles présentées dans le tableau 3.1 page 108, sauf que nous considérons que les images utilisées conjointement pour estimer la profondeur proviennent de deux mises au point différentes, l'une à 2 m et l'autre à 3 m. La figure 3.15 (a) présente les courbes de σ_{BCR} obtenues avec les trois modes de calcul pour des distances comprises entre 1 et 4 m. Les valeurs de l'écart-type en pixel $|\tau|$ des gaussiennes simulant les FEP sont présentées à la figure 3.15 (b) en fonction de la profondeur.

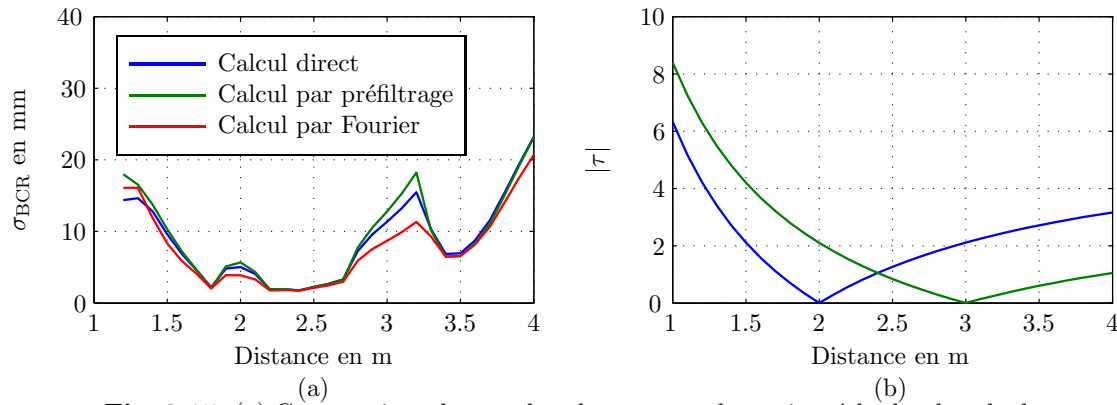


Fig. 3.15: (a) Comparaison des courbes de σ_{BCR} par les trois méthodes de calcul : le calcul direct, le préfiltrage des données et la transformée de Fourier. (b) Écart-type τ en pixel des gaussiennes simulant la FEP pour chaque profondeur et les deux mises au point.

La figure 3.15 montre que l'allure des courbes de σ_{BCR} obtenues avec les trois méthodes de calcul est similaire. L'écart entre les valeurs de σ_{BCR} par le calcul direct et par le préfiltrage des données est inférieur à 5 % sauf au niveau des plans de mise au point où elle

atteint 10%. L'écart est plus important dans le cas d'un calcul par transformée de Fourier, il est en moyenne inférieur à 10 % sauf au niveau des plans de mise au point où il atteint 25%.

Le temps de calcul pour une profondeur dans le cas direct est de 37 s, par préfiltrage il est de 1 s, et par transformée de Fourier il est de 30 ms, avec un processeur CPU Intel Core 2 Duo de fréquence 3GHz. A nouveau, la technique de préfiltrage est la méthode qui permet d'avoir les résultats les plus proches du calcul direct, avec un temps de calcul réduit.

Concernant l'allure des courbes, nous observons que contrairement au cas SA-DFD, les valeurs de σ_{BCR} ne présentent pas de divergence verticale aux abords des plans de mise au point. En effet, l'utilisation de deux images obtenues avec des plans de mise au point différents permet d'avoir toujours au moins l'une des deux FEP dont la taille est supérieure à celle du pixel. Cependant, la précision diminue localement au niveau des deux plans de mise au point car dans ces régions, l'une des deux images n'est plus informative, l'estimation de profondeur repose alors sur une seule image et elle est donc moins précise. La figure 3.15, en comparaison avec la figure 3.1, montre que la précision obtenue dans le cas MA-DFD est bien meilleure que dans le cas SA-DFD, pour des imageurs les mêmes caractéristiques.

3.4.2 Optimisation du rapport des flous entre les deux images

Dans [Rajagopalan et Chaudhuri, 1998], la BCR est utilisée pour optimiser le rapport des flous entre deux images utilisées pour estimer la profondeur. Dans cette section, l'objectif est de comparer les résultats obtenus dans cette référence, qui utilise un modèle de scène de type ARMA gaussien et le calcul de BCR que nous proposons. Les auteurs montrent qu'outre la solution triviale correspondant à avoir une image nette et une image floue, il existe une valeur du rapport du flou entre les deux images qui minimise la BCR. La figure 3.16 montre la comparaison entre les résultats obtenus avec notre calcul de BCR et ceux obtenus dans [Rajagopalan et Chaudhuri, 1998]. La figure 3.16 (a) présente les résultats obtenus en utilisant le calcul de BCR par préfiltrage. Les FEP simulées sont des gaussiennes et comme dans l'article de Rajagopalan et Chaudhuri, l'image de référence a un flou d'écart type $\sigma_1 = 3$. Nous faisons varier le rapport des flous en modifiant l'ouverture de l'imageur. La figure 3.16 (b) correspond à la figure obtenue dans [Rajagopalan et Chaudhuri, 1998].

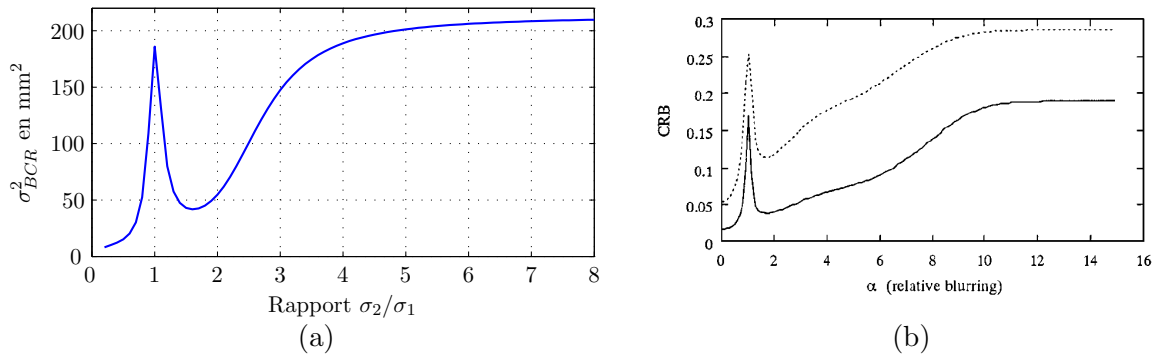


Fig. 3.16: (a) Variation de σ_{BCR} en fonction du rapport des flous de défocalisation entre deux images σ_2/σ_1 . (b) Courbes de $\sigma_{BCR}^2 = \text{CRB}$ obtenues dans [Rajagopalan et Chaudhuri, 1998].

Dans les figures 3.16 (a) et (b), σ_{BCR} est minimale pour un rapport entre les flous tendant vers 0, ce qui revient à avoir une image nette et une image floue. Ensuite, la précision présente un pic au niveau d'un rapport des flous égal à 1. Ceci n'est pas surprenant car alors les

deux images ont le même flou ce qui est la configuration la moins informative. Après ce pic, σ_{BCR} diminue et atteint une valeur minimale pour un rapport de flou autour de 1.7. Après cette valeur, σ_{BCR} augmente et atteint une divergence horizontale. La figure 3.16 montre que l'allure générale de σ_{BCR} obtenue par notre calcul est similaire à celle présentée dans [Rajagopalan et Chaudhuri, 1998]. Notamment, ils estiment la position du minimum à 1.73, ce qui correspond à la valeur que nous obtenons malgré un modèle de scène différent.

3.4.3 Études sur des images simulées

L'objectif de cette section est de comparer les valeurs de σ_{BCR} avec l'écart-type mesuré sur des images simulées à l'aide de l'algorithme MA-GLDFD décrit dans le chapitre 2 à la page 75. Les FEP d'un imageur sont simulées à l'aide des formules d'optique de Fourier présentées dans l'annexe B pour l'imageur du tableau 3.1 avec une mise au point à 2 m et à 3 m. Les distances considérées sont 1.5 m, 1.8 m, 2 m, 2.5 m, 3.5 m et 4 m. Pour chaque distance, une famille de FEP potentielle est générée dans un intervalle de 10 cm de largeur, avec un pas de 2 mm.

Dans un premier temps, pour chaque distance, 140 fenêtres de données sont générées à partir d'une scène respectant l'a priori gaussien et avec un niveau de bruit tel que $\alpha = 0.001$. La figure 3.17 (a) présente les valeurs d'écart-type mesurées à partir des résultats obtenus avec l'algorithme MA-GLDFD en comparaison avec les valeurs de σ_{BCR} .

Puis pour chaque distance, 140 fenêtres sont simulées à partir d'extraits de scènes naturelles avec un bruit blanc d'écart-type 0.01 sachant que les scènes sont normalisées. Les scènes utilisées sont présentées à l'annexe F. La figure 3.17 (b) présente les valeurs d'écart-type empirique σ_e obtenues pour chaque profondeur à partir de ces fenêtres, en enlevant 5% des valeurs extrémales. Ces résultats sont comparés aux valeurs de σ_{BCR} calculées pour chaque profondeur avec la valeur moyenne des α estimés par l'algorithme MA-GLDFD pour l'ensemble des fenêtres.

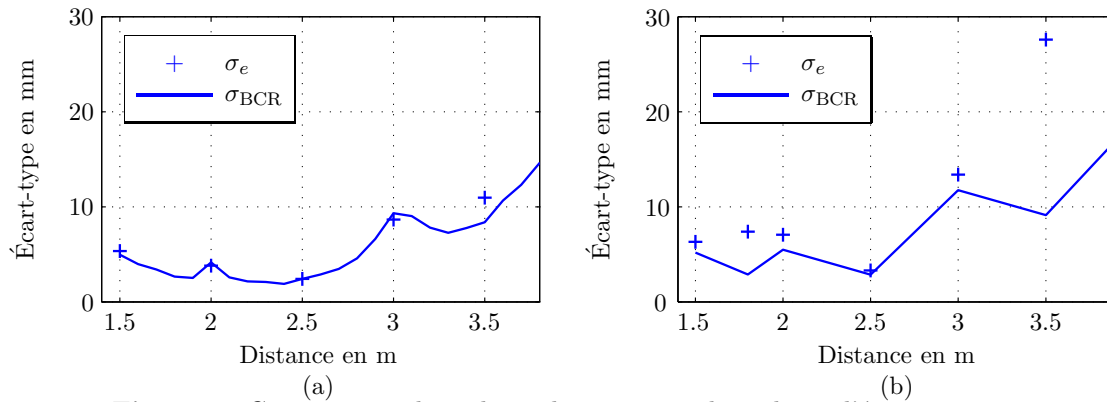


Fig. 3.17: Comparaison des valeurs de σ_{BCR} avec les valeurs d'écart-type empirique obtenues : (a) sur des données simulées à partir de scènes vérifiant l'a priori gaussien. (b) sur des données simulées à partir de scènes naturelles.

Pour les images générées à partir de scènes vérifiant l'a priori gaussien, les valeurs d'écart-type mesurées suivent bien les variations de σ_{BCR} . Nous pouvons observer que le biais est négligeable devant l'écart-type mesuré. Pour les fenêtres générées à partir de scènes naturelles, les courbes de σ_{BCR} et l'écart-type empirique sont en bon accord jusqu'à 3 m, au delà l'écart entre les deux valeurs est plus important. Nous pouvons constater que le point à 3.5 m est associé à une valeur de biais plus importante que pour les autres profondeurs, de

l'ordre de 10 mm. À cette profondeur, l'une des deux images traitées comporte un niveau de flou important, ce qui peut expliquer ce comportement différent par rapport aux autres profondeurs.

3.4.4 Bilan

Dans cette section, nous avons montré comment généraliser le calcul de la BCR développé dans le cas SA-DFD au cas MA-DFD. Ce cas est pour nous une étape intermédiaire dans le but de développer le calcul de la BCR dans le cas SAC-DFD qui nous intéresse particulièrement. C'est pourquoi nous n'avons pas effectué de validation expérimentale afin de vérifier la cohérence entre la prévision de la BCR et la précision expérimentale sur des images réelles. Cependant, nous avons montré que les résultats obtenus avec notre calcul de BCR étaient cohérents avec ceux obtenus dans la référence [Rajagopalan et Chaudhuri, 1998]. De plus, nous avons vérifié que sur des images simulées, les valeurs d'écart-type empirique suivent les variations théoriques de σ_{BCR} .

3.5 Étude de la performance d'estimation de profondeur dans le cas SAC-DFD

Cette section est consacrée à l'étude de performance d'estimation de profondeur dans le cas de l'utilisation d'une optique chromatique. Le formalisme utilisé dans le chapitre 2, qui consiste à modéliser la scène en utilisant la décomposition en luminance et chrominances est repris ici pour le calcul de la BCR. Trois méthodes de calcul différentes sont présentées et comparées sur un exemple simple, impliquant des FEP gaussiennes. Puis les valeurs de σ_{BCR} sont comparées aux valeurs de l'écart-type empirique obtenues sur des images simulées et réelles avec l'algorithme SAC-GLDFD, présenté dans le chapitre 2.

3.5.1 Calcul de la BCR dans le cas SAC-DFD

3.5.1.1 Calcul direct

Comme dans le cas MA-DFD, la BCR chromatique peut être calculée en utilisant le formalisme utilisé dans le chapitre 2, page 77. La matrice de précision des données s'écrit alors

$$Q_p = \frac{1}{\sigma_b^2} (I - H_C(H_C^t H_C + \alpha D_C^t D_C)^{-1} H_C^t)$$

avec

$$H_C = \begin{bmatrix} aH_R & bH_R & cH_R \\ dH_V & eH_V & fH_V \\ gH_B & hH_B & iH_B \end{bmatrix}. \quad (3.86)$$

Les coefficients a à i correspondent aux coefficients de la matrice de passage de la décomposition LC1C2 à la décomposition RVB (cf. section 2.4.3 page 77) définie par :

$$\begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_V \\ \mathbf{x}_B \end{bmatrix} = M \begin{bmatrix} \mathbf{x}_L \\ \mathbf{x}_{C_1} \\ \mathbf{x}_{C_2} \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} \mathbf{x}_L \\ \mathbf{x}_{C_1} \\ \mathbf{x}_{C_2} \end{bmatrix}. \quad (3.87)$$

De plus la matrice D_C est définie par :

$$D_C = \begin{bmatrix} \sqrt{\mu_c} D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{bmatrix}, \quad (3.88)$$

où D est la concaténation de matrice de convolution associée aux dérivées horizontales et verticales. Le calcul direct consiste à construire la matrice H_C à partir des FEP de chaque canal, ainsi que la matrice D_C et d'en déduire la matrice Q_p et la valeur de la BCR en utilisant (3.23).

3.5.1.2 Calcul de la matrice de covariance par préfiltrage des données

Un préfiltrage des données permet également de calculer la BCR par la matrice de covariance. Ainsi, l'image de chaque canal c préfiltrée s'écrit :

$$\tilde{\mathbf{y}}_c = w * \mathbf{y}_c = w * h_c * \mathbf{x}_c + w * \mathbf{b}_c. \quad (3.89)$$

En remplaçant \mathbf{x}_c par la somme pondérée des \mathbf{x}_Y , \mathbf{x}_{C_1} et \mathbf{x}_{C_2} l'équation (3.89) devient :

$$\tilde{\mathbf{y}}_R = w * \mathbf{y}_R = w * h_R * [a\mathbf{x}_L + b\mathbf{x}_{C_1} + c\mathbf{x}_{C_2}] + w * \mathbf{b}_R, \quad (3.90)$$

$$\tilde{\mathbf{y}}_V = w * \mathbf{y}_V = w * h_V * [d\mathbf{x}_L + e\mathbf{x}_{C_1} + f\mathbf{x}_{C_2}] + w * \mathbf{b}_V, \quad (3.91)$$

$$\tilde{\mathbf{y}}_B = w * \mathbf{y}_B = w * h_B * [g\mathbf{x}_L + h\mathbf{x}_{C_1} + i\mathbf{x}_{C_2}] + w * \mathbf{b}_B. \quad (3.92)$$

La matrice de covariance a alors la forme suivante :

$$R_C = \begin{pmatrix} R_{RR} & R_{RV} & R_{RB} \\ R_{VR} & R_{VV} & R_{VB} \\ R_{BR} & R_{BV} & R_{BB} \end{pmatrix}, \quad (3.93)$$

où chaque terme R_{cc} est construit par transformation des matrices de corrélation :

$$C_{RR} = C(\tilde{\mathbf{y}}_R) = C(h_R) * [a^2 C(\tilde{\mathbf{x}}_L) + b^2 C(\tilde{\mathbf{x}}_{C_1}) + c^2 C(\tilde{\mathbf{x}}_{C_2})] + \sigma_b^2 C(w), \quad (3.94)$$

$$C_{VV} = C(\tilde{\mathbf{y}}_V) = C(h_V) * [d^2 C(\tilde{\mathbf{x}}_L) + e^2 C(\tilde{\mathbf{x}}_{C_1}) + f^2 C(\tilde{\mathbf{x}}_{C_2})] + \sigma_b^2 C(w), \quad (3.95)$$

$$C_{BB} = C(\tilde{\mathbf{y}}_B) = C(h_B) * [g^2 C(\tilde{\mathbf{x}}_L) + h^2 C(\tilde{\mathbf{x}}_{C_1}) + i^2 C(\tilde{\mathbf{x}}_{C_2})] + \sigma_b^2 C(w). \quad (3.96)$$

Or d'après les résultats de la section 3.3.1.2 les covariances des vecteurs $\tilde{\mathbf{x}}_L$, $\tilde{\mathbf{x}}_{C_1}$ et $\tilde{\mathbf{x}}_{C_2}$ sont :

$$C(\tilde{\mathbf{x}}_L) = \frac{\sigma_x^2}{4\mu_c} w, \quad C(\tilde{\mathbf{x}}_{C_1}) = \frac{\sigma_x^2}{4} w, \quad \text{et} \quad C(\tilde{\mathbf{x}}_{C_2}) = \frac{\sigma_x^2}{4} w. \quad (3.97)$$

Ainsi :

$$C_{RR} = \sigma_x^2 \left(\frac{1}{4} [a^2/\mu_c + b^2 + c^2] C(h_R) * w + \alpha C(w) \right), \quad (3.98)$$

$$C_{VV} = \sigma_x^2 \left(\frac{1}{4} [d^2/\mu_c + e^2 + f^2] C(h_V) * w + \alpha C(w) \right), \quad (3.99)$$

$$C_{BB} = \sigma_x^2 \left(\frac{1}{4} [g^2/\mu_c + h^2 + i^2] C(h_B) * w + \alpha C(w) \right). \quad (3.100)$$

En supposant que les vecteurs \mathbf{x}_L , \mathbf{x}_{C_1} , \mathbf{x}_{C_2} , \mathbf{b}_R , \mathbf{b}_V et \mathbf{b}_B sont décorrélés, les termes croisés de la covariance se calculent à partir des matrices de corrélation suivantes :

$$C_{RV} = \frac{\sigma_x^2}{4} [ad/\mu_c + be + cf] C(h_R, h_V) * w, \quad (3.101)$$

$$C_{RB} = \frac{\sigma_x^2}{4} [ag/\mu_c + bh + ci] C(h_R, h_B) * w, \quad (3.102)$$

$$C_{VB} = \frac{\sigma_x^2}{4} [gd/\mu_c + he + if] C(h_V, h_B) * w. \quad (3.103)$$

3.5.1.3 Calcul en utilisant une décomposition en Fourier

Comme dans les cas SA-DFD et MA-DFD, la décomposition des matrices dans le domaine de Fourier permet également d'accélérer le calcul de la BCR. Cependant, comme dans le cas MA-DFD cela demande une réorganisation des données. L'objectif est tout d'abord de calculer l'inverse du terme : $H_C^t H_C + \alpha D_C^t D_C$. Pour cela, la matrice de convolution de chaque canal H_c est diagonalisée suivant la forme :

$$H_c^t H_c = W^* |\tilde{H}_c|^2 W, \quad (3.104)$$

où W est la matrice unitaire associée à la matrice de Fourier d'ordre N et $*$ est l'opérateur adjoint (transposée conjuguée) [Idier, 2001]. De même la matrice D_C est décomposée sous la forme :

$$D_C^t D_C = W_3^* |\Delta_{\tilde{D}}|^2 W_3 = W_3^* \begin{bmatrix} \mu_c |\tilde{D}|^2 & 0 & 0 \\ 0 & |\tilde{D}|^2 & 0 \\ 0 & 0 & |\tilde{D}|^2 \end{bmatrix} W_3. \quad (3.105)$$

avec

$$W_3 = \begin{bmatrix} W & 0 & 0 \\ 0 & W & 0 \\ 0 & 0 & W \end{bmatrix}. \quad (3.106)$$

et où $|\tilde{D}|^2$ a été définie à la section 3.3.1.3. Cette décomposition permet d'écrire :

$$H_C^t H_C = W_3^* |\Delta_{\tilde{H}}|^2 W_3, \quad (3.107)$$

avec :

$$\Delta_{\tilde{H}} = \begin{bmatrix} a\tilde{H}_R & b\tilde{H}_R & c\tilde{H}_R \\ d\tilde{H}_V & e\tilde{H}_V & f\tilde{H}_V \\ g\tilde{H}_B & h\tilde{H}_B & i\tilde{H}_B \end{bmatrix}. \quad (3.108)$$

La matrice de précision Q_p s'écrit donc sous la forme : $Q_p = W_3^* \tilde{Q}_p W_3$. Avec :

$$\tilde{Q}_p = \frac{1}{\sigma_b^2} (I - \Delta_{\tilde{H}} (|\Delta_{\tilde{H}}|^2 + \alpha |\Delta_{\tilde{D}}|^2)^{-1} \Delta_{\tilde{H}}^*). \quad (3.109)$$

Dans un premier temps, il s'agit donc d'inverser la matrice $|\Delta_{\tilde{H}}|^2 + \alpha |\Delta_{\tilde{D}}|^2$ que l'on peut écrire sous la forme :

$$|\Delta_{\tilde{H}}|^2 + \alpha |\Delta_{\tilde{D}}|^2 = \begin{bmatrix} O_p & S_p & T_p \\ S_p & V_p & F_p \\ T_p & F_p & Z_p \end{bmatrix}, \quad (3.110)$$

sachant que toutes les matrices $O_p, S_p, T_p, V_p, F_p, Z_p$ sont des matrices diagonales. Il est possible de transformer $(|\Delta_{\tilde{H}}|^2 + \alpha |\Delta_{\tilde{D}}|^2)$ en une matrice diagonale par bloc \tilde{G}_p en regroupant les valeurs par groupe de taille 3×3 correspondant à des fréquences identiques. Ainsi \tilde{G}_p peut s'écrire

$$\tilde{G}_p = \begin{bmatrix} \tilde{G}_p^r(1) & 0 & 0 & \dots & 0 \\ 0 & \tilde{G}_p^r(2) & 0 & \dots & 0 \\ & & \tilde{G}_p^r(m) \dots & & \\ 0 & \dots & & \tilde{G}_p^r(N) \end{bmatrix} \text{ avec } \tilde{G}_p^r(m) = \begin{bmatrix} O(m, m) & S(m, m) & T(m, m) \\ S(m, m) & V(m, m) & F(m, m) \\ T(m, m) & F(m, m) & Z(m, m) \end{bmatrix}. \quad (3.111)$$

\tilde{G}_p est donc une matrice diagonale par bloc qui s'inverse facilement. Avec cette nouvelle représentation des données, la matrice Q_p s'écrit :

$$\tilde{Q}_p = \begin{bmatrix} \Delta_{\tilde{Q}_1} & 0 & 0 & \dots & 0 \\ 0 & \Delta_{\tilde{Q}_2} & 0 & \dots & 0 \\ & & & \dots & \\ 0 & \dots & & & \Delta_{\tilde{Q}_N} \end{bmatrix} \quad (3.112)$$

Avec $\Delta_{\tilde{Q}_m} = (I - \Delta_{\tilde{H}_m}(\tilde{G}_m)^{-1}\Delta_{\tilde{H}_m}^t)$ et

$$\Delta_{\tilde{H}_m} = \begin{bmatrix} a\tilde{H}_R(m, m) & b\tilde{H}_R(m, m) & c\tilde{H}_R(m, m) \\ d\tilde{H}_V(m, m) & e\tilde{H}_V(m, m) & f\tilde{H}_V(m, m) \\ g\tilde{H}_B(m, m) & h\tilde{H}_B(m, m) & i\tilde{H}_B(m, m) \end{bmatrix}. \quad (3.113)$$

Pour calculer la BCR chaque bloc de la matrice Q_p est inversé séparément sauf le premier bloc qui correspond à la fréquence nulle et qui n'est pas inversible. Finalement l'information de Fisher s'écrit :

$$\text{IF}(p) = \frac{1}{2} \sum_{m \neq m_0}^N \text{tr} \left(\Delta_{\tilde{Q}_m}^{-1} \frac{d\Delta_{\tilde{Q}_m}}{dp} \Delta_{\tilde{Q}_m}^{-1} \frac{d\Delta_{\tilde{Q}_m}}{dp} \right). \quad (3.114)$$

3.5.1.4 Exemple

Les trois méthodes de calcul précédentes sont comparées sur un exemple d'imageur chromatique simulé en utilisant des FEP gaussiennes. Les caractéristiques de cet imageur sont présentées dans le tableau 3.5.

f en mm	R 25.2 ; V 25 ; B 24.9
$F/\#$	3
p_x	5 μm
Mise au point en m	R ∞ ; V 2.6 ; B 1.8
ρ	0.25
Taille des fenêtres	21 \times 21
Taille des FEP	21 \times 21
α	0.001
Pas dérivation	1 mm
μ_c	0.05

Tableau 3.5: Paramètres de l'imageur chromatique simulé.

La figure 3.18 montre les résultats des calculs de σ_{BCR} avec les trois méthodes. L'allure de σ_{BCR} est identique selon les trois méthodes de calcul. Cependant il existe un écart moyen de l'ordre de 2% entre la valeur de σ_{BCR} calculée par préfiltrage des données et par le calcul direct et un écart de 8% entre le calcul par transformée de Fourier et le calcul direct. Sous **Matlab**, le calcul direct pour une profondeur prend 3 min, le calcul par préfiltrage 2.5 s et le calcul par transformée de Fourier 80 ms, avec un processeur CPU Intel Core 2 Duo de fréquence 3GHz. A nouveau, le calcul par préfiltrage est le plus intéressant en termes de temps de calcul et d'écart avec le calcul direct. Concernant l'allure de la courbe de σ_{BCR} elle-même, comme pour le cas MA-DFD, il n'existe plus de divergence verticale mais simplement une augmentation locale de σ_{BCR} au niveau des plans de mise au point. Une étude plus approfondie de la BCR chromatique et sa comparaison avec le cas MA-DFD est faite dans la section suivante.

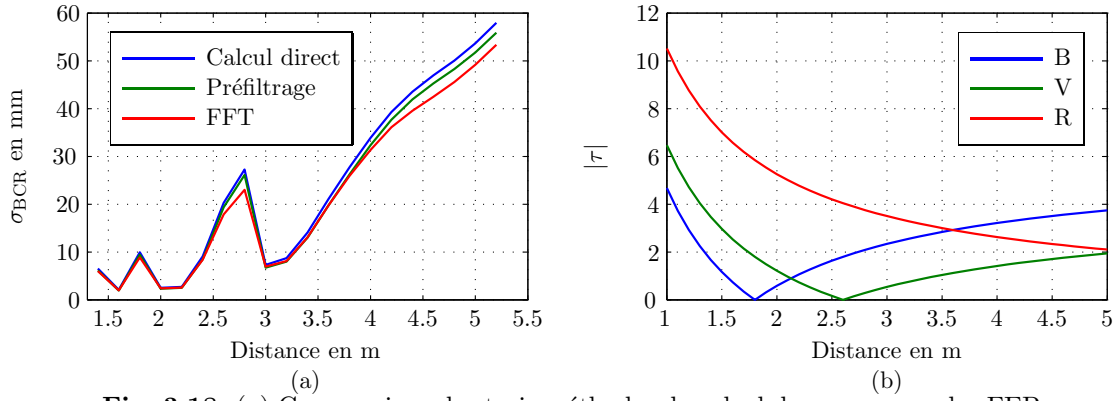


Fig. 3.18: (a) Comparaison des trois méthodes de calcul de σ_{BCR} pour des FEP gaussiennes : calcul direct, préfiltrage des données et décomposition en base de Fourier. (b) Variations de l'écart-type de la gaussienne représentant la FEP de chaque canal RVB en fonction de la profondeur.

3.5.2 Influence du modèle de scène chromatique sur la BCR

Selon nous, l'estimation de profondeur chromatique peut-être interprétée comme un cas intermédiaire entre l'estimation de profondeur de type MA-DFD, à partir de trois images d'une même scène avec un détecteur noir et blanc, et la combinaison de l'estimation de profondeur de type SA-DFD à partir de trois images de scènes décorréées. L'objectif de cette section est de retrouver cette interprétation sur les courbes de σ_{BCR} en faisant varier le paramètre μ_c dans le modèle chromatique et en comparant les performances obtenues avec le cas MA-DFD et la combinaison d'estimation de type *SADFD* sur trois images décorréées.

Comme discuté à la page 80, selon [Condat, 2009], le cas $\mu_c = 1$ correspond au cas où la corrélation des trois scènes RVB n'est pas prise en compte dans l'estimation de la profondeur et le cas $\mu_c = 0$ au cas d'une scène en noir et blanc, où les trois scènes RVB sont identiques. Les courbes de σ_{BCR} obtenues pour un imageur simulé avec des FEP simulées à l'aide des formules de l'annexe B sont donc calculées pour trois valeurs de μ_c , un μ_c très faible, $\mu_c = 1$ et $\mu_c = 0.05$ et représentée à la figure 3.19(a).

Pour commencer, nous comparons les courbes de performance obtenues lorsque le paramètre $\mu_c = 1$ avec le cas de l'estimation de profondeur à partir de trois images de scènes décorréées. Dans ce cas, σ_{BCR} se calcule à partir de la relation :

$$\text{IF}(p) = \sum_{i=R,V,B} \text{IF}_i(p). \quad (3.115)$$

où IF_i correspond à l'information de Fisher dans le cas SA-DFD. La figure 3.19(b) présente l'écart entre les courbes de σ_{BCR} chromatique obtenue pour $\mu_c = 1$ et celle obtenue par la formule (3.115). L'écart entre ces deux courbes est négligeable, ce qui confirme l'analyse de Condat : le cas $\mu_c = 1$ correspond au cas où les trois images RVB sont totalement décorréées.

Par ailleurs, d'après Condat, lorsque μ_c tend vers 0, les trois scènes RVB sont complètement corrélées, ce qui correspond à une scène en noir et blanc. La BCR chromatique doit donc tendre vers celle obtenue dans le cas MA-DFD. Pour comparer les courbes de σ_{BCR} dans ces deux cas, il faut que le rapport signal à bruit soit identique. Or, dans le cas MA-DFD ou SAC-DFD, la définition du paramètre de régularisation α n'est pas la même. En effet, soit σ_x^2 la variance des gradients de l'image en noir et blanc \mathbf{x} . Dans la décomposition en luminance et chrominance, la variance de la luminance vaut alors $3\sigma_x^2$ puisque

$\mathbf{x}_L = 1/\sqrt{3}(\mathbf{x}_R + \mathbf{x}_V + \mathbf{x}_B) = \sqrt{3}\mathbf{x}$ et que la scène est en noir et blanc. Or dans le modèle chromatique, la variance de la luminance est modélisée par $\sigma_x'^2/\mu_c$, où σ_x' correspond à la variance des gradients des chrominances. Nous en déduisons donc $\sigma_x'^2 = 3\mu_c\sigma_x^2$. Ainsi, en posant $\alpha_{\text{MA-DFD}} = \sigma_b^2/\sigma_x^2$ et $\alpha_{\text{SAC-DFD}} = \sigma_b^2/\sigma_x'^2$, le paramètre de régularisation dans le cas chromatique est alors $\alpha_{\text{SAC-DFD}} = \alpha_{\text{MA-DFD}}/(3\mu_c)$. Les courbes de la figure 3.19 (c) sont tracées en utilisant cette relation. Le cas chromatique avec μ_c tendant vers 0 se rapproche alors bien du cas MA-DFD. Cet exemple illustre le fait qu'une faible valeur de μ_c modélise une scène en noir et blanc.

Finalement, cette étude illustre que le cas chromatique est un cas intermédiaire entre le cas de scènes totalement corrélées et le cas où les images sont traitées comme si elles étaient totalement décorrélées.

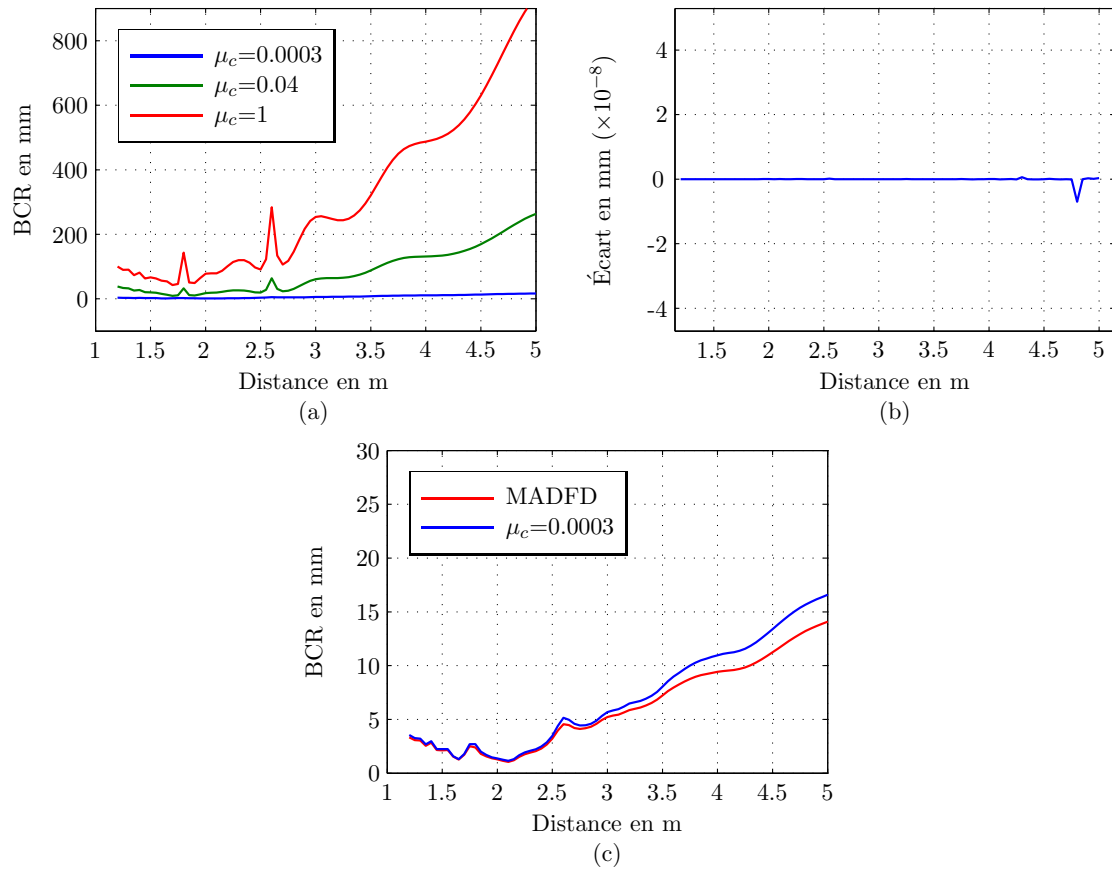


Fig. 3.19: (a) Variation de σ_{BCR} dans le cas SAC-DFD pour trois valeurs de μ_c . (b) Ecart entre les valeurs de σ_{BCR} calculées pour trois images sans considérer une corrélation entre les scènes en utilisant la formule de l'équation (3.115) et la formule de σ_{BCR} pour $\mu_c = 1$. (c) comparaison des valeurs de σ_{BCR} calculées dans le cas SAC-DFDs avec $\mu_c = 1$ et $\mu_c \rightarrow 0$ avec respectivement le cas de scènes considérées comme décorrélées et le cas MA-DFD.

3.5.3 Comparaison de l'écart-type empirique avec σ_{BCR} sur des images simulées

L'objectif de cette section est de comparer les valeurs de σ_{BCR} dans le cas chromatique avec l'écart-type des résultats d'estimation de profondeur obtenus avec l'algorithme SAC-GLDFD. Pour cela, un imageur chromatique est modélisé avec des plans de mise au point

RVB respectivement à l'infini, 2.6 m et 1.8 m en simulant les FEP en utilisant les formules d'optique de Fourier présentées dans l'annexe B. Les FEP de chaque canal RVB sont simulées avec des longueurs d'onde valant respectivement 470, 550 et 640 nm, qui correspondent aux maxima des spectres B, V et R d'un détecteur couleur.

3.5.3.1 Comportement théorique

Le tableau 3.6 présente les caractéristiques de l'imageur simulé. La figure 3.20(a) montre les performances théoriques d'estimation de profondeur pour cet imageur. La figure 3.20 (b) présente la variation de la taille géométrique de la FEP $|\epsilon|$ des trois canaux RVB.

f en mm	R 25.2 ; V 25 ; B 24.9
$F/\#$	3
p_x	5 μm
Mise au point en m	R ∞ ; V 2.6 ; B 1.8
Taille des fenêtres	21 \times 21
Taille des FEP	21 \times 21
α	0.001
Pas dérivation	1 mm

Tableau 3.6: Paramètres de l'imageur chromatique simulé.

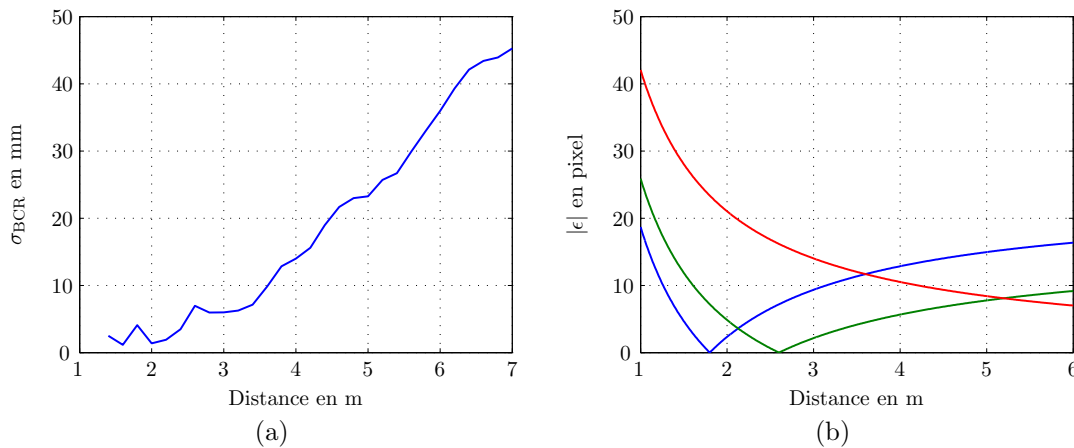


Fig. 3.20: (a) Valeurs de σ_{BCR} en fonction de la profondeur pour le système optique chromatique dont les paramètres sont décrits dans le tableau 3.6. (b) Variation de $|\epsilon|$ avec la profondeur pour les trois plans de mise au point correspondant aux canaux RVB.

Nous observons une augmentation presque régulière de l'incertitude théorique avec, comme dans le cas MA-DFD, une légère augmentation du paramètre σ_{BCR} au niveau des plans de mise au point des canaux RVB.

3.5.3.2 Étude sur des données simulées

Les FEP des trois canaux RVB de l'imageur décrit dans le tableau 3.20 sont simulées dans un intervalle centré autour des profondeurs 2 m, 3 m, 4 m, 5 m, et 5.5 m. Les valeurs de σ_{BCR} sont ensuite comparées aux valeurs d'écart-type empirique mesurées à l'aide de

l'algorithme SAC-GLDFD, successivement sur des images simulées à partir de scènes vérifiant l'a priori gaussien, puis sur des scènes naturelles.

Dans un premier temps, pour chaque profondeur, un ensemble de 140 fenêtres de taille 21×21 pixels sont simulées à partir de scènes respectant l'a priori gaussien générées afin d'avoir $\alpha = 0.001$. L'algorithme SAC-GLDFD est ensuite appliqué sur ces fenêtres. La figure 3.21(a) présente les valeurs d'écart-type des résultats obtenus pour chaque profondeur en comparaison avec les valeurs de σ_{BCR} . Ensuite, les images sont simulées à partir de scènes naturelles. Pour chaque profondeur, un ensemble de 140 fenêtres de taille 21×21 pixels est extrait de 10 scènes naturelles convoluées à la FEP correspondant à cette profondeur, les scènes utilisées étant présentées dans l'annexe F. Un bruit blanc gaussien d'écart-type 0.01 est ajouté au résultat, sachant que les scènes sont normalisées à 1. La figure 3.21(b) présente l'écart-type empirique obtenu en comparaison avec les valeurs de σ_{BCR} pour les valeurs moyennes de α estimées par l'algorithme SAC-GLDFD sur les 140 fenêtres.

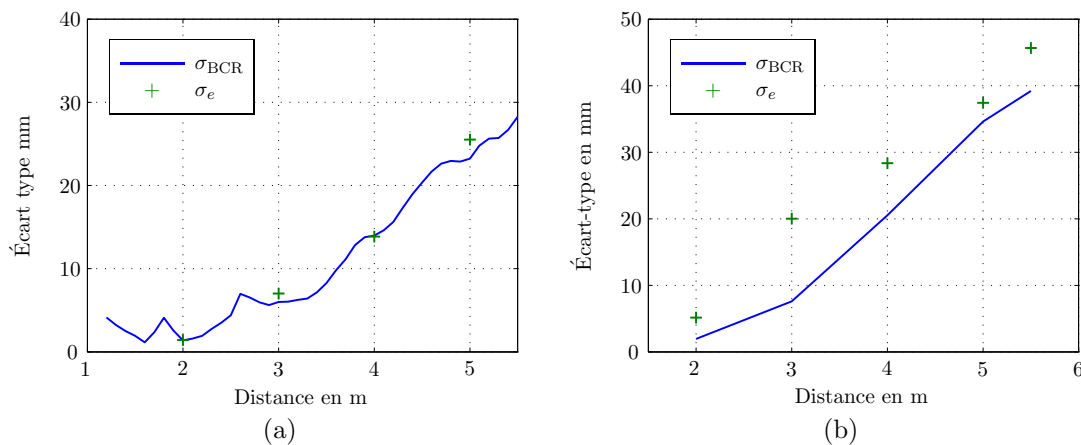


Fig. 3.21: Comparaison des valeurs de σ_{BCR} et des valeurs d'écart-type empirique obtenues avec l'algorithme SAC-GLDFD pour des images simulées : (a) à partir de scènes vérifiant l'a priori gaussien, (b) à partir de scènes naturelles.

Dans les deux types d'images simulées, les valeurs de σ_{BCR} sont proches des valeurs d'écart-type mesurées expérimentalement. Comme dans les cas SA-DFD et MA-DFD, les valeurs d'écart-type empirique mesurées sur des images produites à partir de scènes vérifiant l'a priori gaussien sont plus proches des valeurs de σ_{BCR} que lorsque les images sont produites à partir de scènes naturelles. Notons que le biais est à nouveau négligeable par rapport à l'écart-type empirique dans le cas de scènes vérifiant l'a priori gaussien et qu'il augmente lorsque les scènes sont naturelles.

3.5.3.3 Étude sur des données réelles

À l'aide du prototype d'imageur chromatique présenté dans le chapitre 2, il est possible de mesurer expérimentalement l'écart-type empirique d'estimation de profondeur et de le comparer aux valeurs de σ_{BCR} calculées à partir des FEP simulées en utilisant les formules d'optique de Fourier présentées dans l'annexe B. En effet, nous disposons des plans des lentilles et de leurs caractéristiques. La figure 3.22(a) présente le diamètre géométrique des FEP $|\epsilon|$ pour les trois canaux RVB de la mise au point utilisée. La figure 3.22(c) présente les résultats d'estimation de profondeur obtenus sur la cible « Parterre » dans la section 2.4.5 sur 100 fenêtres de taille 21×21 pixels. La figure 3.22(b) présente les valeurs de σ_{BCR}

obtenues en prenant α , pour chaque profondeur, égal à la valeur moyenne des valeurs de α estimées par l'algorithme SAC-GLDFD sur l'ensemble des fenêtres traitées.

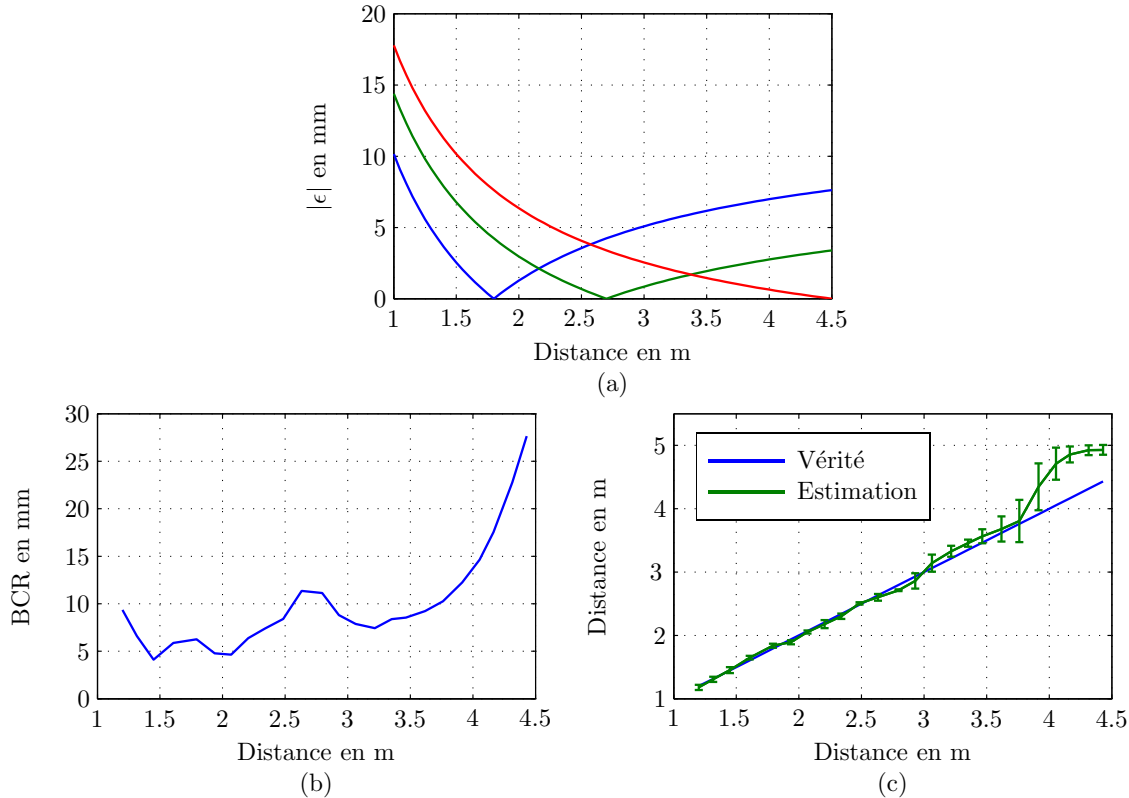


Fig. 3.22: (a) Variation du diamètre géométrique de la FEP $|\epsilon|$ en pixel avec la profondeur pour les trois canaux RVB. (b) Valeurs de σ_{BCR} . (c) Résultats d'estimation de profondeur expérimentaux sur la cible « Parterre ».

Sur la courbe de σ_{BCR} , l'écart-type vaut environ 5 cm et augmente légèrement entre 1 et 3 m, puis une augmentation plus rapide apparaît à partir de 3.5 m. Cette augmentation de l'incertitude peut s'expliquer par le fait qu'à partir d'environ 4 m, la FEP du canal rouge est inférieure à un pixel, donc à partir de cette distance, l'estimation repose sur les deux autres canaux V et B qui sont flous et ne permettent donc pas une estimation de profondeur très précise. Sur les résultats d'estimation de profondeur expérimentaux, présentés à la figure 3.22(c), l'écart-type augmente effectivement à partir de 3 m. Vers 4 m l'estimation devient fortement biaisée, donc le comportement ne peut plus s'expliquer par la BCR. Cependant, nous pouvons toutefois observer que la BCR prédit une dégradation des performances vers 4 m et que cette dégradation est observée expérimentalement, comme le montrent les résultats de la figure 3.22(c), mais cette dégradation est plus grave car elle s'accompagne de l'augmentation du biais.

3.5.4 Bilan

Nous avons montré comment calculer la BCR dans le cas d'un imageur chromatique. Ce calcul reprend la modélisation de la scène proposée dans le chapitre 2 utilisant la décomposition en luminance et chrominances. Nous avons montré que le cas de l'estimation de profondeur chromatique est intermédiaire entre le cas de l'acquisition de trois images à partir de scènes totalement décorréliées et le cas MA-DFD où trois images sont produites

par la même scène. Ce résultat indique que l'utilisation du chromatisme donnera des résultats moins précis que si l'on utilise une technique de type MA-DFD, mais en contrepartie l'acquisition est moins contraignante car les trois images sont obtenues simultanément. Les valeurs de σ_{BCR} et d'écart-type empirique obtenu à partir des résultats d'estimation de profondeur par l'algorithme SAC-GLDFD sur des images simulées sont proche. La comparaison des valeurs de σ_{BCR} vis-à-vis de l'écart-type empirique sur des données réelles obtenues à partir de l'imageur chromatique dont nous disposons n'est pas évidente à cause des biais expérimentaux, mais la BCR prédit effectivement que l'estimation de profondeur sera moins précise à partir de 4 m ce qui est observé expérimentalement.

3.6 Études complémentaires

Les courbes de performance théorique étudiées dans les sections précédentes correspondent à une modélisation simplifiée. Tout d'abord, nous avons considéré uniquement la variation de la BCR avec la profondeur, or Q_p dépend de deux autres paramètres qui sont le rapport signal à bruit et la variance du bruit, dont l'estimation est menée conjointement à l'estimation de la profondeur. De plus, nous avons tracé des courbes de σ_{BCR} qui modélisent la variation de la BCR avec la profondeur pour un rapport signal à bruit constant. Or en pratique, lorsque la scène est réculée par rapport à l'imageur, le facteur d'échelle introduit une modification des gradients de la scène, autrement dit une variation du paramètre σ_x^2 et donc du paramètre α . Enfin, nous avons négligé l'effet du schéma Bayer, donc du sous-échantillonnage des données pour l'estimation de profondeur. Dans cette section, nous revenons sur ces différentes simplifications.

3.6.1 Variation de la BCR avec les autres paramètres de la matrice de précision

Dans les sections précédentes, le seul paramètre considéré dans le calcul de la BCR était la profondeur. Or, nous avons vu dans l'équation (3.14) page 100 que la matrice de précision $Q_y(\theta)$ dépend également de deux autres paramètres qui sont le paramètre $\alpha = \sigma_b^2/\sigma_x^2$, considéré comme l'inverse d'un rapport signal à bruit, et l'écart-type du bruit σ_b . Dans le cas de l'estimation de plusieurs paramètres θ , la matrice de Fisher IF est une matrice $n \times n$ définie par :

$$IF_{k,l}(\theta) = \frac{1}{2} \text{tr} \left(Q_y(\theta)^{-1} \frac{\partial Q_y(\theta)}{\partial \theta_k} Q_y(\theta)^{-1} \frac{\partial Q_y(\theta)}{\partial \theta_l} \right). \quad (3.116)$$

Nous définissons ici le vecteur de paramètre $\theta = \{p, \alpha, 1/\sigma_b^2\}$. Notons que nous prenons pour troisième paramètre l'inverse de la variance du bruit car cela permet d'obtenir une expression simple de la dérivée partielle de Q_θ vis-à-vis de ce paramètre. Ainsi, les dérivées partielles de Q_θ par rapport aux autres paramètres que la profondeur s'écrivent :

$$\frac{\partial Q_\theta}{\partial \alpha} = \frac{1}{\sigma_b^2} H_p (H_p^t H_p + \alpha D^t D)^{-1} D^t D (H_p^t H_p + \alpha D^t D)^{-1} H_p^t \quad (3.117)$$

$$\frac{\partial Q_\theta}{\partial (1/\sigma_b^2)} = I - H_p (H_p^t H_p + \alpha D^t D)^{-1} H_p^t. \quad (3.118)$$

À l'aide des formules (3.117) et (3.118), nous construisons, pour chaque profondeur, la matrice IF de taille 3×3 définie dans l'équation (3.116) par le calcul direct. Cette matrice est ensuite inversée et la valeur de BCR associée à l'estimation de profondeur, prenant en compte l'estimation des deux autres paramètres, correspond alors à la racine carrée du

premier terme de la diagonale de cette matrice. La figure 3.23 montre l'effet de l'estimation des deux autres paramètres sur la BCR par rapport aux courbes de BCR en considérant uniquement l'estimation de la profondeur, pour deux valeurs de α .

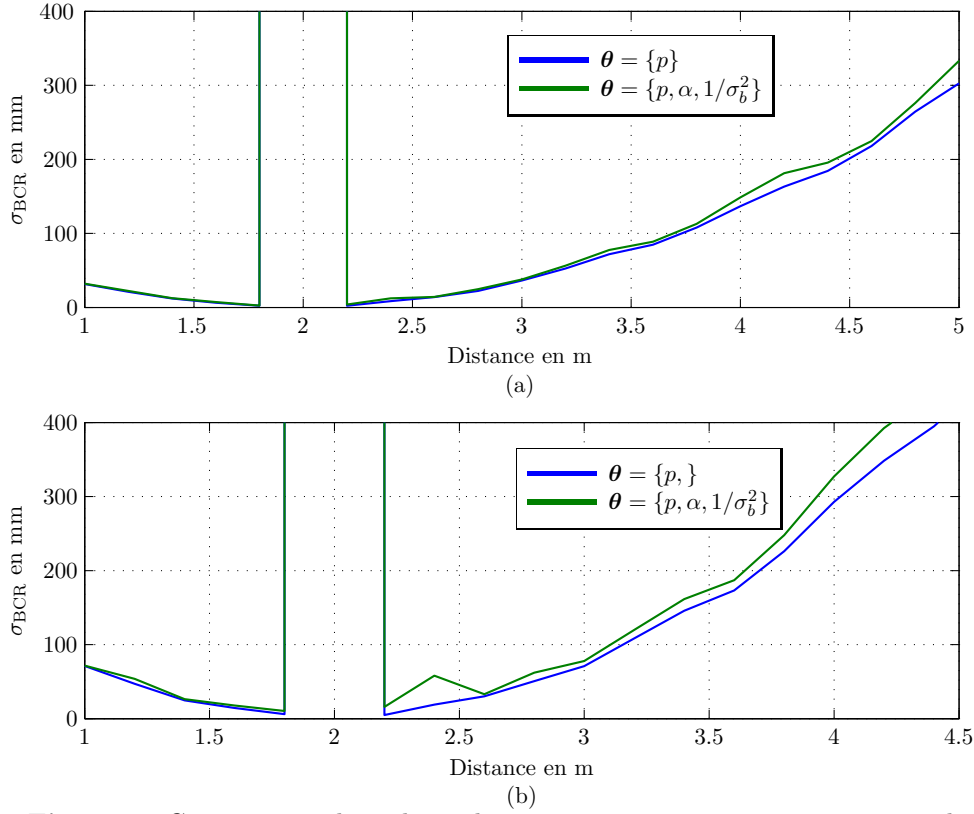


Fig. 3.23: Comparaison des valeurs de σ_{BCR} en tenant compte uniquement de l'estimation de p et en tenant compte de l'estimation de α et σ_b^2 , pour deux valeurs de α : (a) $\alpha = 0.001$ et (b) $\alpha = 0.01$.

Dans la figure 3.23(a) $\alpha = 0.001$, donc le rapport signal à bruit est élevé. Dans ce cas l'estimation de p en tenant compte des erreurs d'estimation des autres paramètres diffère peu du cas où p est le seul paramètre estimé, l'écart entre les deux courbes de performance est de l'ordre de 10%. Dans la figure 3.23(b), $\alpha = 0.01$, donc le rapport signal à bruit est plus faible et l'écart précédent augmente. Il est alors de l'ordre de 15 %.

Nous avons montré comment tenir compte des erreurs d'estimation des paramètres de rapport signal à bruit et de variance, dans la BCR associée à l'estimation de la profondeur. Cependant, cette correction ne semble pas entraîner une modification significative de la courbe de σ_{BCR} , ce qui motive a posteriori notre choix de ne pas tenir compte de cet effet dans les études comparatives menées précédemment.

3.6.2 Modélisation du recul de la scène

Dans ce chapitre les courbes de σ_{BCR} sont calculées pour différentes profondeurs avec une valeur de α fixe, ce qui revient à considérer un rapport signal à bruit constant avec la profondeur. Cependant, pour modéliser l'estimation de profondeur avec une scène fixe et un imageur se rapprochant de cette scène, le rapport signal à bruit ne peut plus être considéré comme constant. En effet, à cause du grandissement variable avec la profondeur,

l'échantillonnage de la scène idéale nette varie, ce qui introduit une variation du paramètre σ_x^2 et donc du paramètre α . Dans cette section nous modélisons la variation du paramètre α avec la profondeur et observons l'effet de cette variation sur les valeurs de σ_{BCR} dans les cas SA-DFD, MA-DFD et SAC-DFD.

L'image défocalisée apparaît sur le détecteur à l'échelle de l'image géométrique idéale et sa taille dépend de la profondeur p et de la distance lentille/détecteur d_{det} . En effet, le grandissement est alors défini par $\gamma = \frac{d_{\text{det}}}{p}$.

Soient $Y_0(m, n, p_0)$ et $Y(m, n, p)$ les images géométriques idéales de la cible lorsqu'elle est placée respectivement à p_0 et p . Le facteur de grandissement relatif entre ces deux images peut être défini par $\gamma' = \frac{\gamma(p_0)}{\gamma(p)} = \frac{p_0}{p}$ ce qui implique $Y(m, n, p) = Y_0(\frac{m}{\gamma'}, \frac{n}{\gamma'}, p_0)$. Ainsi :

$$\frac{\partial Y}{\partial m}(m, n, p) = \frac{1}{\gamma'} \frac{\partial Y_0}{\partial m} \left(\frac{m}{\gamma'}, \frac{n}{\gamma'}, p_0 \right) \quad (3.119)$$

Soient $\sigma_x^2(p)$ la variance des gradients de Y et $\sigma_{x_0}^2$ celle de Y_0 , en considérant que la relation (3.119) est conservée par discrétisation des images, σ_x^2 et $\sigma_{x_0}^2$ vérifient :

$$\sigma_x^2 = \gamma'^{-2} \sigma_{x_0}^2 = \left(\frac{p}{p_0} \right)^2 \sigma_{x_0}^2(p_0). \quad (3.120)$$

La même relation est obtenue dans la direction n . Ainsi, d'après ce modèle, la variance des gradients augmente avec la profondeur. La simulation présentée à la figure 3.24 illustre ce phénomène. La scène initiale, modélisée par le signal 1D de la figure 3.24(a), est supposée être placée à 1 m, puis cette dernière est reculée progressivement de l'imageur, ce qui entraîne un rééchantillonnage d'un facteur γ' . Ce rééchantillonnage est effectué avec la fonction `resample` de `Matlab` et la variance du gradient horizontal du nouveau signal est calculée pour différentes profondeurs. La figure 3.24(b) représente la variation de l'écart-type σ_x obtenu en fonction de la profondeur p en comparaison avec la formule (3.120).

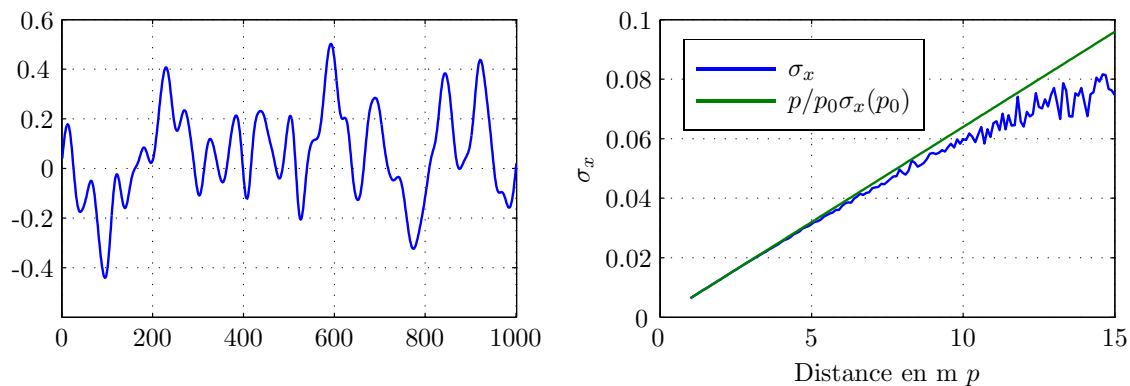


Fig. 3.24: (a) Signal aléatoire représentant la scène à 1 m. (b) Variation de l'écart-type des gradients avec la profondeur.

Pour des profondeurs inférieures à environ 7 m, l'écart-type des gradients augmente conformément à la formule (3.120). Cependant pour des profondeurs plus importantes, la valeur mesurée est instable. En effet, pour des profondeurs importantes, le sous-échantillonnage de l'image introduit du repliement ce qui perturbe la variance des gradients de la scène.

En négligeant ce repliement, la variation du paramètre α avec la profondeur peut donc être modélisée par :

$$\alpha(p) = \frac{p_0^2}{p^2} \alpha(p_0) \quad (3.121)$$

La figure 3.25 montre les courbes de σ_{BCR} pour les cas SA-DFD, MA-DFD et SAC-DFD, en tenant compte ou non de l'effet de la variation de α avec la profondeur avec $p_0 = 1\text{m}$ et $\alpha(p_0) = 0.001$. Pour chaque profondeur les matrices $Q_{p-\delta}$, Q_p et $Q_{p+\delta}$ sont calculées en faisant varier à la fois H et α avec la profondeur.

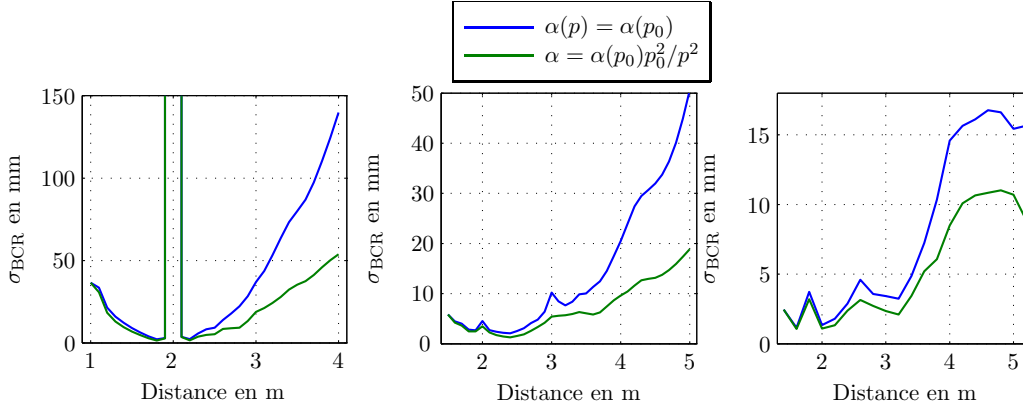


Fig. 3.25: Effet de la modélisation du recul de la scène sur les courbes de σ_{BCR} dans les cas : (a) SA-DFD, (b) MA-DFD et (c) SAC-DFD.

Cette figure illustre que la modélisation du recul de la scène fait décroître significativement les courbes de σ_{BCR} mais l'allure générale de ces courbes reste identique.

3.6.3 Effet du filtre de Bayer

Lorsque l'acquisition est effectuée avec un détecteur couleur possédant un filtre de Bayer, les images R, V et B sont sous-échantillonnées. Comme discuté dans le chapitre 1, cet effet peut être modélisé en enlevant des lignes de la matrice de convolution H correspondant aux pixels manquants. Nous présentons à la figure 3.26 des comparaisons des courbes de σ_{BCR} dans les cas SA-DFD, MA-DFD et SA-DFD, en tenant compte ou non de ce sous-échantillonnage.

Dans les cas SA-DFD et MA-DFD, σ_{BCR} est tout d'abord calculé en considérant que le détecteur est noir et blanc, ce qui n'implique aucun sous-échantillonnage, puis en considérant que l'estimation de profondeur est réalisée uniquement sur les pixels des canaux R, V ou B d'un détecteur couleur. Nous obtenons ainsi les courbes des figures 3.26(a) et (b). Le sous-échantillonnage réduit la précision d'estimation de profondeur, ce qui n'est pas surprenant car il y a une perte d'information due à la diminution du nombre de pixels traités. Puisque le canal vert présente deux fois plus de pixels que les canaux R et B, la précision d'estimation de profondeur est meilleure dans ce canal.

Dans le cas SAC-DFD, σ_{BCR} est calculé dans le cas d'un détecteur tri-CCD qui fournit trois images RVB de même résolution, sans sous-échantillonnage. Puis σ_{BCR} est calculée dans le cas d'un détecteur couleur de type Bayer, où le sous-échantillonnage des canaux RVB est modélisé. A nouveau, σ_{BCR} augmente lorsque le détecteur possède un filtre de Bayer, du fait de la réduction du nombre de données utilisées pour l'estimation de profondeur.

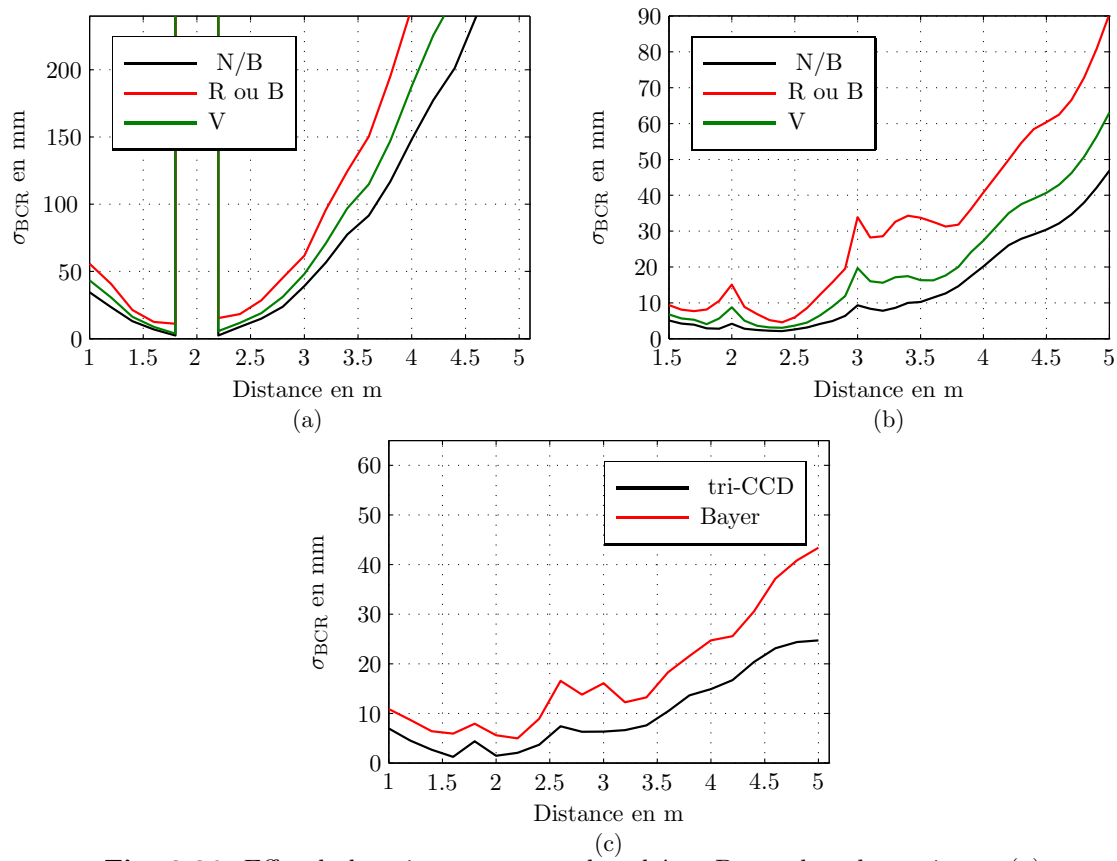


Fig. 3.26: Effet de la prise en compte du schéma Bayer dans les trois cas (a) SA-DFD, (b) SAC-DFD et (c) MA-DFD.

En conclusion, dans les trois cas SA-DFD et MA-DFD et SAC-DFD, la modélisation du schéma Bayer entraîne une augmentation de σ_{BCR} . Cependant, l'allure des courbes de performance est identique.

La prise en compte du sous-échantillonnage permet de modéliser plus finement le processus d'acquisition dans le cas de l'utilisation d'un détecteur couleur de type Bayer. Ce sous-échantillonnage se modélise simplement dans le calcul direct de la BCR, en enlevant des lignes de la matrice de convolution, mais il est plus délicat de le prendre en compte dans le calcul de la BCR par préfiltrage des données. En s'inspirant du travail de [Alleysson *et al.*, 2005], il est possible de tenir compte du sous-échantillonnage dans la diagonalisation des matrices en Fourier, mais cette piste n'a pas été explorée dans cette thèse. Au final, avec les techniques dont nous disposons, la modélisation du sous-échantillonnage lié à la présence d'un filtre de Bayer reste donc coûteuse en temps de calcul. Comme par ailleurs le sous-échantillonnage ne modifie pas l'allure des courbes de σ_{BCR} , nous avons choisi de ne pas tenir compte de cet effet.

3.7 Comparaisons théoriques d'imageurs estimant la profondeur par SA-DFD ou SAC-DFD

Dans les sections précédentes nous avons présenté différentes méthodes de calcul de BCR associées à plusieurs types d'imageur passif monovoie. L'objectif de cette section est de comparer les performances d'estimation de profondeur de différentes solutions optiques et différents traitements. À l'aide des formules de BCR que nous avons développées, il est possible de conduire de multiples comparaisons, en faisant varier les valeurs d'ouvertures, de distances focales et de mises au point des imageurs ainsi que les types de traitement. Nous montrons ici quelques exemples de comparaison de systèmes optiques permettant d'estimer la profondeur à partir d'un détecteur couleur, qui est le cas le plus courant aujourd'hui et qui nous intéresse particulièrement dans le cadre de la conception d'une caméra à capacité 3D couleur. Parmi les différents paramètres optiques, nous choisissons de fixer la distance focale à 25 mm, et l'ouverture à 2.5. De plus, la taille des pixels est fixée à $5\ \mu\text{m}$ et le rapport signal à bruit est tel que $\alpha = 0.001$. Nous faisons ici varier la nature de l'objectif (objectif achromatique ou chromatique) et le type de traitement (SA-DFD ou SAC-DFD) dans le cas d'une même optique chromatique.

3.7.1 Comparaison de deux solutions optiques pour le même détecteur et le même algorithme d'estimation de profondeur

Dans un premier temps, nous comparons les performances d'estimation de profondeur obtenues avec un imageur chromatique et avec un imageur achromatique de même ouverture et de même focale. Dans les deux cas, le détecteur est un détecteur couleur, de type tri-CCD et la technique d'estimation de profondeur est de type SAC-DFD. L'imageur achromatique a son plan de mise au point à 2.6 m et pour l'imageur chromatique les plans de mise au point RVB sont respectivement à 4 m, 2.6 m et 1.8 m.

La figure 3.27 présente les courbes de σ_{BCR} obtenues avec les deux types d'imageurs. Le premier intérêt du système chromatique est qu'il permet d'éviter la divergence verticale qui existe autour du plan de mise au point du système achromatique. Au voisinage de ce plan de mise au point, les performances d'estimation de profondeur des deux imageurs sont assez voisines, même si la précision est meilleure dans le cas d'un objectif chromatique. Cependant, pour des distances éloignées du plan de mise au point, il y a un net gain en précision de l'optique chromatique par rapport à l'optique achromatique. De plus, ces courbes ne modélisent pas l'ambiguïté existant de part et d'autre du plan de mise au point dans le cas achromatique. Ainsi, l'estimation de profondeur n'est possible que d'un côté ou de l'autre de ce plan dans le cas de l'optique achromatique, alors que la solution chromatique n'a pas cette restriction de la plage de profondeur. Cette étude comparative confirme l'intérêt d'utiliser une optique chromatique à la fois en termes de précision et de portée, autrement dit de plage de distances où la profondeur peut être estimée.

3.7.2 Variation de la performance des traitements pour un même imageur

Dans cet exemple, nous comparons les performances d'estimation de profondeur obtenues à l'aide d'un imageur chromatique, en utilisant la méthode SA-DFD séparément sur chaque canal R, V et B ou avec la méthode SAC-DFD en combinant les données des canaux R, V, B. Pour faire cette comparaison, le détecteur est un tri-CCD. La figure 3.28 présente les résultats de σ_{BCR} obtenus dans ces différents cas.

Chaque courbe obtenue avec l'algorithme SAC-GLDFD présente une divergence verticale et une croissance forte de σ_{BCR} de part et d'autre du plan de mise au point. Avec

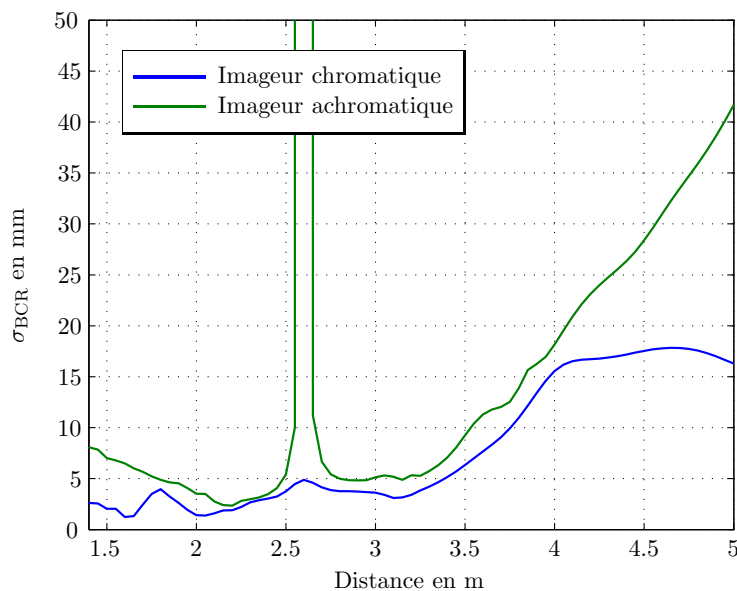


Fig. 3.27: Comparaison des performances théoriques d'un imageur chromatique et d'un imageur achromatique sur le même détecteur couleur tri-CCD. Dans les deux cas, l'estimation de profondeur est réalisée à l'aide de l'algorithme SAC-GLDFD.

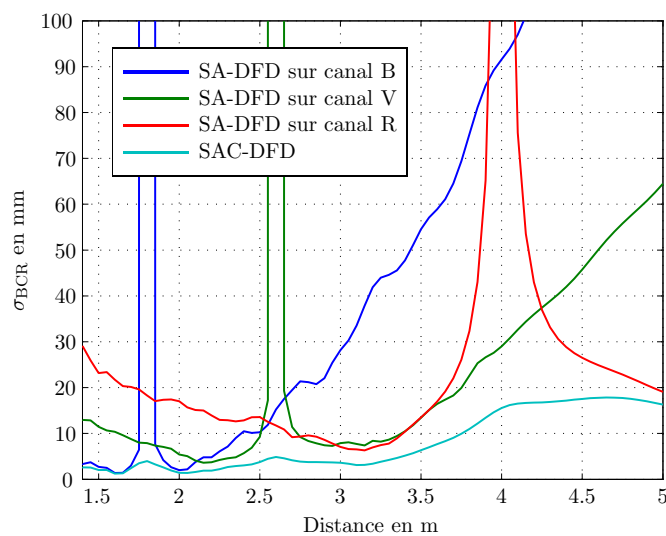


Fig. 3.28: Comparaison de l'estimation de profondeur en utilisant une optique chromatique et chaque canal séparément ou en combinant l'information des trois canaux. Le détecteur est un détecteur tri-CCD.

l'algorithme SAC-GLDFD, la combinaison de l'information contenue dans les trois canaux permet d'éviter les divergences verticales et permet également d'améliorer la précision d'estimation de profondeur. Le gain en précision s'explique notamment par l'utilisation d'un nombre plus important de pixels.

3.8 Conclusion

Dans ce chapitre, nous avons proposé un cadre théorique pour la modélisation de la performance de l'estimation de profondeur à l'aide du flou de défocalisation. Les outils que nous avons développés dans le cas SA-DFD, reposant sur un calcul de la BCR utilisant les modélisations de la scène proposées dans les chapitres 1 et 2, ont également été étendus aux cas MA-DFD et SAC-DFD. Nous avons conduit dans chaque cas des tests expérimentaux sur des données simulées, et lorsque cela était possible sur des images réelles, qui confirment les tendances et les choix indiqués par le modèle théorique.

Dans le cas d'un imageur conventionnel et de l'estimation de profondeur de type SA-DFD, une première observation des courbes de σ_{BCR} a montré que d'une part la profondeur ne peut pas être estimée précisément dans une région proche du plan de mise au point dont la largeur est donnée par la profondeur de champ. D'autre part, la meilleure précision est atteinte juste après cette région. Nous avons ensuite conduit une étude détaillée de la variation de σ_{BCR} en fonction des paramètres optiques afin de déterminer les facteurs prépondérants sur la précision d'estimation de profondeur. Cette étude a conduit à deux comportements différents de la BCR. En effet, pour des distances proches du plan de mise au point, la meilleure précision d'estimation de profondeur est obtenue pour la configuration présentant la plus grande variation de la FEP avec la profondeur alors que pour des distances éloignées du plan de mise au point, c'est la taille de la FEP qui influence le plus la précision. Ceci favorise alors des FEP de petites tailles, quitte à diminuer la variation de la FEP avec la profondeur. Ce résultat est nouveau car il va à l'encontre de l'intuition qui consiste à vouloir augmenter la sensibilité de la taille de la FEP avec la profondeur pour améliorer la précision. Ces résultats sont confirmés dans ce chapitre par les validations que nous avons menées sur des données simulées et réelles.

Dans les cas MA-DFD et SAC-DFD, nous avons observé que l'utilisation de plusieurs images permet d'éviter les divergences verticales observées dans la BCR mono-image et de diminuer significativement les valeurs de σ_{BCR} par rapport au cas SA-DFD.

Ainsi, en termes de précision absolue, le cas MA-DFD apporte la meilleure précision et le cas SA-DFD présente le moins de contraintes expérimentales. Le cas SAC-DFD offre selon nous le meilleur compromis entre les contraintes de réalisation expérimentale et la précision théorique, car il nécessite simplement de placer une optique chromatique devant un détecteur couleur. Nous avons donc choisi d'utiliser le calcul de la BCR pour la co-conception d'un imageur chromatique à capacité 3D et le résultat de cette co-conception est présentée dans le chapitre suivant.

Co-conception d'un imageur à capacité 3D

Dans les chapitres précédents nous avons développé un algorithme d'estimation de profondeur basé sur l'estimation locale du flou de défocalisation. Puis nous avons validé expérimentalement une nouvelle solution optique favorisant l'estimation de profondeur qui consiste à utiliser un imageur chromatique. Nous avons ensuite développé une modélisation permettant de prévoir la précision théorique pour différentes familles d'imageur monovoie passif, en particulier pour le cas d'un imageur chromatique dont nous avons pu illustrer l'intérêt en termes de précision et de portée d'estimation de profondeur. Notre objectif à présent est d'utiliser ces outils pour concevoir conjointement un imageur chromatique à capacité 3D. Dans ce chapitre, nous présentons une démarche de co-conception que nous avons établie en collaboration avec le Département d'Optique Théorique et Appliquée de l'Onera (DOTA). À partir d'un cahier des charges caractérisant typiquement la portée, la précision de l'imageur et les conditions d'utilisation, nous proposons un ensemble d'étapes permettant d'optimiser un imageur à capacité 3D, en utilisant entre autres le modèle de performance décrit dans le chapitre 3. Le procédé de co-conception ainsi décrit fait l'objet de la demande de dépôt de brevet n° FR 12.02774 déposée le 17 octobre 2012. Cette démarche est mise en œuvre concrètement dans ce chapitre pour la réalisation d'un imageur chromatique co-conçu, appelé CAM3D. Nous présentons les différentes étapes de conception que nous avons utilisées afin d'obtenir son architecture précise. Cet imageur a ensuite été réalisé au DOTA puis nous l'avons calibré afin de caractériser la précision d'estimation expérimentale de profondeur mesurée dans la plage de profondeur définie dans le cahier des charges. Cette première expérience de co-conception d'un imageur à capacité 3D a permis d'illustrer l'intérêt de la démarche de conception conjointe.

ORGANISATION DU CHAPITRE

Pour commencer nous présentons la problématique liée à la conception d'un imageur à capacité 3D en comparaison avec la conception classique. La principale difficulté est qu'il faut alors prédimensionner l'imageur en fonction de nouvelles consignes du cahier des charges portant notamment sur une précision et une portée d'estimation de profondeur. L'enjeu est alors de définir les paramètres qui caractérisent au premier ordre la performance de ce type d'imageur et nous montrons, à l'aide d'un exemple concret, que les positions des plans de mise au point RVB sont des paramètres pertinents pour caractériser cette performance.

La suite de ce chapitre est ensuite dédiée à la conception de l'imageur chromatique co-conçu CAM3D. Nous présentons le cahier des charges et les critères de conception que nous avons utilisés, avant de présenter les résultats d'un dimensionnement préliminaire des caractéristiques de l'imageur. Une première architecture optique a ensuite été optimisée suivant nos critères de performances pour aboutir aux caractéristiques finales de l'optique chromatique. Cette optique a été réalisée et nous comparons les résultats d'estimation de profondeur que nous avons obtenus en l'associant à deux détecteurs possédant des tailles de pixel différentes. Dans la fin de ce chapitre, nous revenons sur les choix de conception et les améliorations que nous pouvons apporter suite à cette première expérience de conception conjointe.

4.1 Introduction

4.1.1 Problématique

Dans ce chapitre, notre objectif est de concevoir conjointement un imageur chromatique à capacité 3D passif reposant sur l'estimation du flou de défocalisation. Comme illustré à la figure 4.1 cet imageur doit être capable de produire une image RVB ainsi qu'une carte de profondeur de la scène imagée.

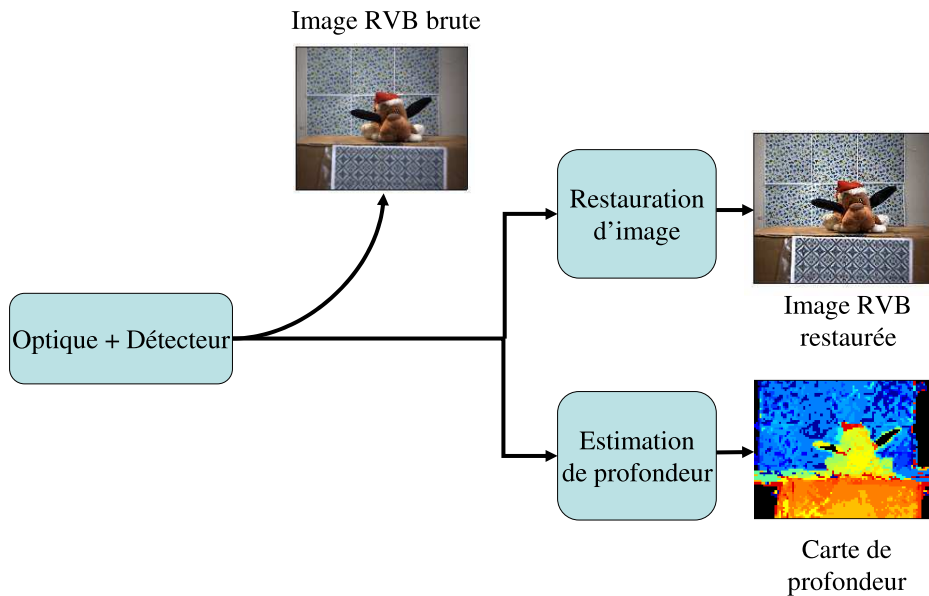


Fig. 4.1: Principe de l'imageur à capacité 3D recherché. Les illustrations proviennent du prototype d'imageur chromatique présenté dans le chapitre 2.

L'imageur est constitué d'une optique chromatique et d'un détecteur qui permettent de produire une image couleur. Cette image est en réalité constituée de trois images R, V et B correspondant chacune à l'image de la scène dans une bande spectrale particulière. Ces images sont traitées par deux modules de traitement en parallèle. Le premier a pour objectif de produire une carte de profondeur et le second restaure une image en couleur de grande profondeur de champ, à partir des trois images RVB possédant des plans de mise au point différents.

Pour concevoir un imageur du type de celui présenté à la figure 4.1, il est nécessaire de mettre en place une démarche de co-conception, c'est-à-dire de prendre en compte les paramètres liés au traitement en plus de ceux liés à l'optique, pour optimiser la performance globale de l'imageur, caractérisée par la précision d'estimation de profondeur et la capacité à produire une image de bonne qualité. Nous nous écartons ainsi de la conception classique qui consiste à optimiser uniquement les paramètres optiques suivant un critère de qualité image. Il n'est cependant pas inutile de rappeler comment se déroule la conception classique d'un imageur. Le point de départ est l'exploitation du cahier des charges qui définit typiquement le champ d'observation, la résolution de l'imageur et les conditions d'illuminations. Ces données permettent alors de faire un dimensionnement préliminaire des caractéristiques principales de l'imageur afin de définir la distance focale, l'ouverture, la dimension du détecteur et la taille des pixels. À partir du dimensionnement préliminaire, l'opticien propose alors une architecture, formée d'un certain nombre de dispositifs optiques, qui lui paraît susceptible de respecter le cahier des charges. Cette architecture est modélisée grâce à un logiciel de conception optique qui prend en entrée les paramètres des lentilles tels que leur position, leur courbure, leur épaisseur et la nature du verre. Ce logiciel permet d'évaluer, entre autres, les aberrations de l'imageur telles que l'aberration sphérique, le chromatisme, l'astigmatisme, la coma, la courbure ou la distorsion. Il permet également de corriger ces aberrations à l'aide d'algorithmes d'optimisation reposant sur des critères caractérisant la qualité image, en général par rapport au rayon de la FEP ou par rapport au front d'onde.

Pour concevoir notre imageur, nous proposons de conserver la démarche de conception classique, à savoir l'exploitation d'un cahier des charges pour faire un dimensionnement préliminaire de l'imageur, le choix d'une architecture et une optimisation de cette architecture. Cependant, nous introduisons dans cette démarche deux nouveaux critères de conception, qui caractérisent d'une part la précision d'estimation de profondeur, en s'appuyant sur le modèle de performance de l'estimation de profondeur développé au chapitre 3 et d'autre part la capacité à produire une image RVB nette sur une large plage de profondeurs. L'un des points délicats de cette démarche est l'étape de dimensionnement préliminaire. En effet, le cahier des charges d'un imageur à capacité 3D contient, en plus des contraintes classiques portant sur le champ d'observation et la résolution de l'image, des contraintes sur la portée de l'estimation de profondeur et la précision de cette estimation. La question qui se pose à présent est comment exploiter ces données afin d'en déduire les paramètres principaux de l'imageur chromatique, permettant à l'opticien de faire le choix d'une architecture adaptée? Pour cette étape, il est important de définir les paramètres optiques qui gouvernent, au premier ordre, la performance d'un imageur chromatique à capacité 3D en terme de précision et de profondeur de champ après restauration. Pour ce faire, nous considérons dans la section suivante l'optimisation de la performance de l'imageur conçu à l'IOGS.

4.1.2 Optimisation de la précision de l'imageur conçu à l'IOGS

Nous reprenons ici le prototype d'imageur chromatique conçu à l'IOGS qui a été présenté au chapitre 2, en essayant de modifier son optique pour améliorer la performance d'estimation de profondeur. Pour cela, nous disposons d'un seul degré de liberté qui est la position du détecteur. Dans un premier temps, nous avons modifié cette position dans le logiciel de conception optique **Zemax** qui modélise cet imageur. Pour plusieurs positions du détecteur, nous avons calculé les courbes de précision théorique, caractérisée par l'écart-type σ_{BCR} obtenu par le calcul de la borne de Cramère Rao (BCR) défini dans le chapitre 3. Le paramètre σ_{BCR} est obtenu ici en utilisant les FEP simulées par **Zemax** et le modèle

de performance correspondant au cas SAC-DFD, décrit dans la section 3.5 du chapitre 3. Nous avons ainsi obtenu une position du détecteur qui améliore nettement les performances théoriques d'estimation de profondeur par rapport à la configuration initiale. Cette nouvelle mise au point a ensuite été reproduite expérimentalement et nous avons à nouveau évalué la performance d'estimation de profondeur expérimentale de l'imageur, pour cette configuration optimisée.

La figure 4.2 présente la comparaison des performances théoriques et expérimentales de l'imageur chromatique pour les deux positions des plans de mise au point. Les figures (a), (c) et (e) correspondent aux performances relatives à la position initiale du détecteur utilisée dans le chapitre 2, et les figures (b), (d) et (f) aux performances relatives à la position optimisée du détecteur. La principale conséquence du déplacement du détecteur est le déplacement des plans de mise au point RVB. Nous illustrons cet effet à l'aide des figures 4.2(a) et (b) qui représentent les variations de la taille géométrique de la FEP ϵ , définie dans l'équation (1) page 4, en fonction de la profondeur, pour les deux positions du détecteur. Dans la configuration initiale, les plans de mise au point RVB sont respectivement situés à 4.7, 2.7 et 1.8 m. Pour la position optimale, ils sont situés à 5.5, 3.6 et 2.3 m. Les figures 4.2 (c) et (d) présentent les courbes de σ_{BCR} en fonction de la profondeur pour les deux positions du détecteur et les figures (e) et (f) présentent les résultats expérimentaux d'estimation de profondeur obtenus sur la mire « Fleurs » de la figure 2.6 page 62. La courbe de précision théorique pour la position optimisée du détecteur montre une nette amélioration de la précision aux alentours de 5 m, qui est confirmée par l'évaluation expérimentale.

La première conclusion à tirer de cet exemple est que la performance peut être effectivement améliorée par une modification du système optique issue d'une étude de performance théorique utilisant le modèle que nous avons développé : il s'agit donc d'un premier résultat concret de co-conception.

La deuxième conclusion est que la performance de l'estimation de profondeur est fortement corrélée à la position des plans de mise au point RVB. Dans cet exemple, nous avons modifié leurs positions en jouant sur la position du détecteur. Notons que dans le cas général, la position de ces plans dépend aussi de la valeur du chromatisme longitudinal, défini comme la différence entre les distances focales associées aux longueurs d'ondes des canaux rouge et bleu.

Mais la position des plans de mise au point n'influence pas uniquement la précision d'estimation de profondeur. En effet, comme le montrent les figures 4.2 (a) et (b), elle conditionne aussi les largeurs des FEP sur l'intervalle de profondeur recherché. Or la complexité du problème de restauration d'une image de grande profondeur de champ dépend fortement de ces largeurs. En particulier, pour utiliser une technique de transfert de fréquence comme celle présentée au chapitre 2, l'idéal est d'avoir toujours au moins l'un des canaux net. Nous pouvons constater que dans cet exemple le nouveau réglage va conduire à une baisse de performance en ce qui concerne la restauration de l'image, notamment dans la région de profondeur comprise entre 1.5 et 2 m. En effet, à cause du recul du plan de mise au point du canal B, le transfert de fréquence n'est possible qu'après 2 m pour le réglage optimisé alors qu'il était possible dès 1.8 m dans le réglage initial. Par ailleurs, la largeur des FEP conditionne également la taille de la fenêtre de données traitée par l'algorithme, car cette fenêtre doit contenir la FEP. Or la résolution de la carte de profondeur est liée à la taille de la fenêtre traitée par l'algorithme. La position des plans RVB conditionne donc également la résolution de la carte de profondeur.

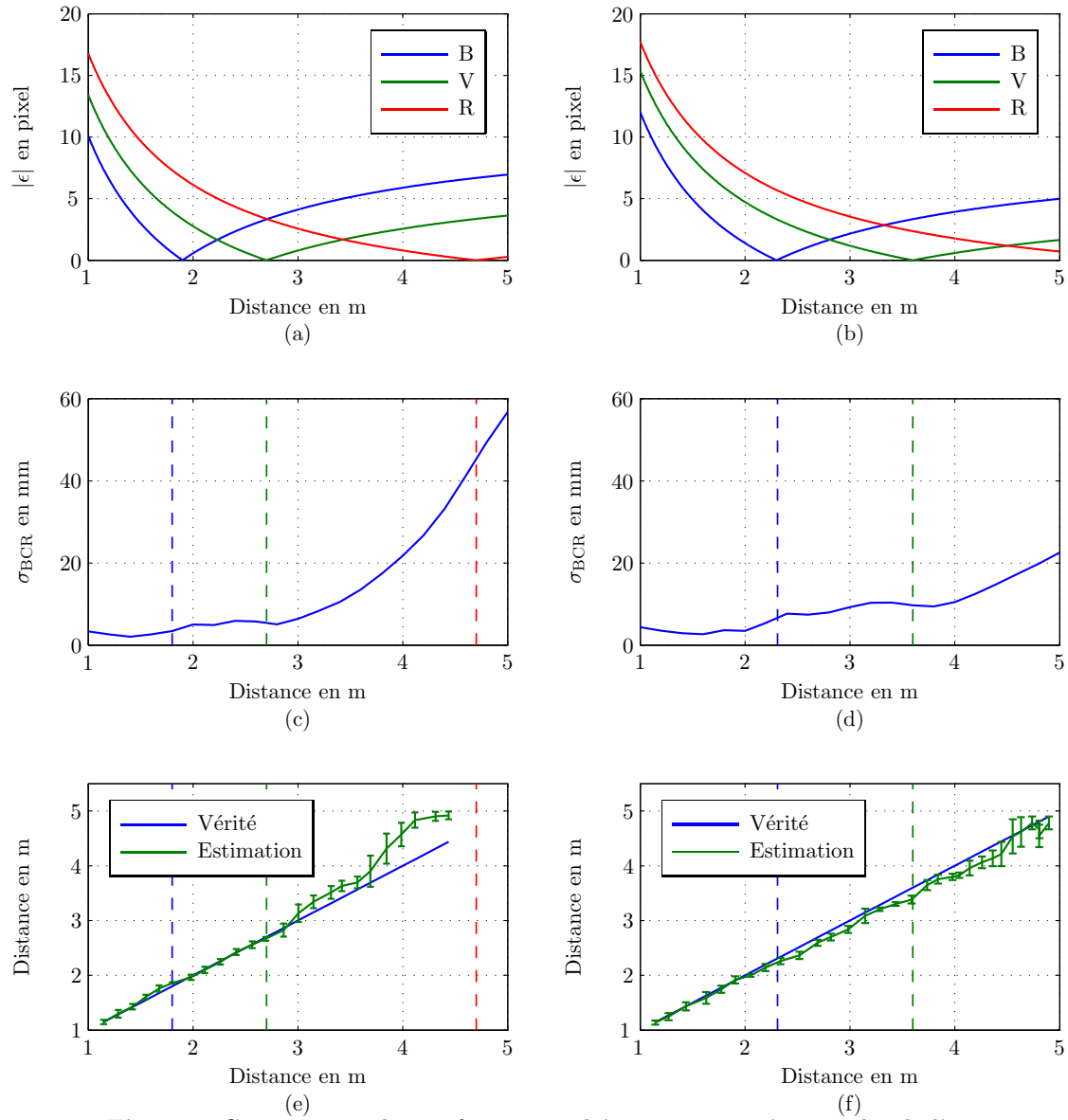


Fig. 4.2: Comparaison des performances théoriques et expérimentales de l'imageur chromatique conçu à l'IOGS, pour deux positions du détecteur différentes. Les figures (a) (c) et (e) correspondent aux résultats théoriques et expérimentaux obtenus avec la position du détecteur utilisée dans le chapitre 2. Les figures (b) (d) et (f) présentent les mêmes résultats mais pour la position optimisée du détecteur. Les droites verticales en pointillés correspondent à la position des plans de mise au point des canaux RVB.

4.1.3 Paramètres optiques choisis pour le dimensionnement préliminaire

Nous choisissons donc de fonder le prédimensionnement de l'imageur chromatique sur la définition des paramètres optiques classiques à savoir la distance focale, l'ouverture et les dimensions du détecteur, auxquels nous proposons d'ajouter la position des plans de mise au point RVB, qui influence la précision, la capacité à produire une image de grande profondeur de champ et la résolution de la carte de profondeur. Ces paramètres permettent de caractériser au premier ordre la performance de l'imageur, par l'intermédiaire du modèle de performance de l'estimation de profondeur développé au chapitre 3 et par un raison-

nement plus qualitatif sur la taille minimale de la FEP en ce qui concerne la qualité de l'image restaurée. Enfin, ces paramètres ont l'avantage d'être facilement interprétables par un opticien pour le guider dans son choix d'architecture.

4.2 Co-conception de l'imageur CAM3D

Dans cette section, nous présentons un exemple de réalisation de la démarche de co-conception que nous proposons, ayant conduit à la réalisation de l'imageur CAM3D. Nous reprenons la démarche classique consistant à exploiter les consignes du cahier des charges pour prédimensionner les paramètres principaux de l'imageur. Puis à partir des paramètres issus du prédimensionnement, une architecture initiale est conçue à l'aide d'un logiciel de conception optique. Cette architecture est alors à son tour optimisée jusqu'à obtenir les caractéristiques finales de l'imageur. La différence avec la conception classique est que nous introduisons de nouveaux critères de conception fondés sur la précision d'estimation de profondeur et sur la capacité à produire une image RVB nette sur une large plage de profondeurs.

4.2.1 Cahier des charges

Le cahier des charges que nous avons adopté dans ce travail résulte de l'étude de définition du projet SPIDER dans lequel s'inscrit cette thèse. Ce projet interne à l'Onera visait à définir des concepts de systèmes perceptifs embarquables permettant la démonstration de capacités de co-conception, avec comme domaine d'application privilégié l'autonomie des micro-drones en environnement urbain. Dans ce cadre et en particulier pour la navigation des micro-drones, l'étude [Champagnat *et al.*, 2010] a souligné l'intérêt de développer un système passif compact capable de fournir une image et une carte de profondeur à une cadence proche de la cadence vidéo. Pour l'évitement d'obstacles, étant donnée la vitesse de vol de ces drones, la portée de l'estimation de profondeur requise est comprise entre 1 et 5 m, avec une résolution permettant de détecter des obstacles de type câbles de hautes tensions, barrières ou échafaudages, soit une résolution de l'ordre de 2 cm à 3 m. Le champ d'observation requis pour cette imagerie est restreint, typiquement de l'ordre de 20 à 30°. En effet, d'autres procédés typiquement fondés sur le flot optique, permettent d'obtenir une information 3D loin de l'axe optique, que nous supposons proche de la direction de déplacement dans le plan image. Les conditions d'emplois en termes d'illumination sont assez exigeantes, le drone pouvant évoluer en intérieur ou en extérieur dans des conditions de lumière faible : le nombre d'ouverture ne doit donc pas être trop élevé. Par ailleurs, nous souhaitons utiliser un minimum de lentilles afin de limiter le coût, la complexité et le poids de l'imageur. Enfin, notre objectif est d'obtenir une estimation de profondeur non biaisée et avec un écart-type de l'ordre de quelques cm pour des profondeurs comprises entre 1 et 5 m.

4.2.2 Définition de critères de conception conjoint

Le premier critère indispensable à la conception d'un imageur à capacité 3D est la précision d'estimation de profondeur. Le travail de modélisation réalisé à la section 3.5 du chapitre 3 permet de calculer cette précision théorique pour une profondeur donnée à partir des FEP de l'imageur chromatique. À partir de ce modèle, il est donc possible de définir un premier critère de performance portant sur la précision d'estimation de profondeur dans une plage de distance donnée. Ce critère peut être défini de différentes manières, nous proposons

d'utiliser la valeur moyenne des valeurs d'écart-type minimal théorique σ_{BCR} calculées entre 1 et 5 m à l'aide des outils développés dans le chapitre 3. Cette moyenne constitue le critère noté C1. Notons que dans ce chapitre nous avons utilisé le calcul de BCR par transformée de Fourier décrit dans le chapitre 3 à la section 3.5.1.3, car il permet une comparaison rapide d'un grand nombre de systèmes optiques. Rappelons que ce calcul ne modélise pas la présence d'un filtre de Bayer sur le détecteur.

Comme discuté dans la section 4.1.1, notre objectif est également d'avoir une image RVB de grande profondeur de champ. Pour cela, une des solutions est d'utiliser un transfert de fréquence, décrit notamment dans la section 2.4.7.2 du chapitre 2. Dans ce cadre, il est alors important d'avoir le plus souvent possible l'un des canaux RVB net. Nous proposons donc un deuxième critère de conception noté C2 qui caractérise la capacité de l'imageur à produire une image de grande profondeur de champ. Ce critère mesure la plage de distance pour laquelle au moins l'une des FEP des trois canaux est inférieure à 1 pixel, si le détecteur est tri-CCD, ou 2 pixels si il est de type Bayer. Dans cette plage de distance nous considérons qu'un transfert de fréquence sera efficace pour restaurer l'image RVB. Concrètement pour calculer le critère C2, nous calculons pour chaque canal c les deux profondeurs $p_{c,\min}$ et $p_{c,\max}$ vérifiant $|\epsilon_c| = p_x$ si le détecteur est de type tri-CCD, ou $|\epsilon_c| = 2p_x$ si le détecteur est de type Bayer, avec $|\epsilon_c|$ défini par l'équation (1) page 4 et p_x la taille du pixel. Nous définissons alors l'ensemble L des profondeurs pour lesquelles un transfert de fréquence est possible par la réunion des segments :

$$L = \bigcup_{c=R,V,B} [p_{c,\min}; p_{c,\max}]. \quad (4.1)$$

Le critère C2 est finalement calculé en déterminant la longueur totale de L en mètre. Il peut être interprété comme une profondeur de champ généralisée de l'image après restauration. Notons que selon la position des plans de mise au point R,V ou B, L peut correspondre à un seul segment de profondeur ou à l'union de plusieurs segments disjoints. De plus, les profondeurs $p_{c,\min}$ ou $p_{c,\max}$ peuvent ne pas appartenir à l'intervalle de profondeur compris entre 1 et 5 m. Dans ce cas, les profondeurs en dehors de ces deux bornes ne sont pas prises en compte dans le calcul de C2.

Notons que d'autres critères caractérisant la qualité de l'image restaurée pourraient être définis en fonction des traitements envisagés pour la restauration de l'image. En l'occurrence, si un traitement de déconvolution tel que celui proposé dans [Soulez et Thiébaud, 2009] est utilisé, il est par exemple possible de définir un critère caractérisant la qualité de l'image reposant sur le rapport signal à bruit de l'image restaurée.

4.2.3 Détermination des paramètres principaux de l'imageur

Dans cette section nous déduisons du cahier des charges un certain nombre de paramètres de l'imageur. Nous commençons par choisir un détecteur ce qui nous amène ensuite à fixer la distance focale, une ouverture et la taille des fenêtres traitées par l'algorithme.

4.2.3.1 Choix du détecteur

Nous disposons de deux détecteurs couleurs de type Bayer de caractéristiques différentes. Le premier est de la marque Stingray, il s'agit du détecteur que nous avons utilisé dans le chapitre 2 avec l'imageur chromatique conçu à l'IOGS. Le deuxième est de la marque Dalsa. Les caractéristiques de ces deux détecteurs sont présentées dans le tableau 4.1.

Détecteur	Dalsa	Stingray
Nombre pixels $N_{px,v} \times N_{px,h}$	1024×1400	2056×2452
Taille des pixels (p_x) en μm	7.4	3.45
Dimension du détecteur(mm \times mm)	7.6×10.4	7.1×8.5
Fill factor	60%	Non connu
Rendement quantique (entre 500 et 700 nm)	>50%	entre 30 et 50 %

Tableau 4.1: Caractéristiques principales des deux détecteurs couleurs Stingray et Dalsa.

D'un côté le détecteur Dalsa présente des pixels de taille plus importante et un rendement quantique plus important que le détecteur Stingray, ce qui est un avantage du point de vue du rapport signal à bruit, pour un même temps d'acquisition. De plus, la fréquence d'échantillonnage est plus faible que celle du détecteur Stingray, il est donc plus facile du point de vue conception optique de faire en sorte que la résolution de l'imageur soit limitée par le détecteur. En contrepartie, cette fréquence d'échantillonnage faible conduit à un risque plus important de repliement spectra au niveau des plans de mise au point. Ce phénomène se traduit par l'ajout des fréquences de la scène supérieures à la fréquence de Nyquist au spectre du signal échantillonné, ce qui perturbe le spectre de l'image au niveau des plans de mise au point. De plus, le détecteur est un peu plus grand que le détecteur Stingray, ce qui est un peu plus contraignant du point de vue de la conception optique.

D'un autre côté, le détecteur Stingray présente des pixels plus petits que le détecteur Dalsa. La fréquence d'échantillonnage est alors plus grande, donc les données et les FEP seront échantillonnées avec plus de points et le risque de repliement spectral est moins important. En contrepartie la taille plus petite des pixels et le rendement quantique plus faible vont diminuer le flux lumineux intégré par les pixels du détecteur, et donc réduire le rapport signal à bruit, en supposant les autres sources de bruit, telles que le bruit de lecture, sont équivalentes et le temps d'acquisition identique pour les deux détecteurs. En outre, pour une résolution donnée de la carte de profondeur, le nombre de données à traiter est plus important avec le détecteur Stingray ce qui augmente le coût de calcul. Enfin, le détecteur Stingray est plus petit que le détecteur Dalsa, donc pour une résolution de la carte de profondeur donnée, le nombre de points de profondeur mesurés sera plus faible avec le détecteur Stingray qu'avec le détecteur Dalsa.

Pour la conception de CAM3D, il nous a paru intéressant de comparer les performances de ces deux détecteurs sur une même optique afin d'étudier expérimentalement l'influence du pas des pixels sur la performance d'estimation de profondeur.

4.2.3.2 Distance focale

Pour un imageur classique, le choix de la distance focale dépend de la taille du détecteur et du champ de l'imageur demandé dans le cahier des charges. Par ailleurs, l'imageur chromatique ne possède pas une unique distance focale car celle-ci varie avec la longueur d'onde. Cependant, la distance focale va varier faiblement autour d'une grandeur que l'on peut définir à partir d'une taille de détecteur et d'un champ et associer par exemple au canal vert. Les valeurs de distance focale des autres canaux seront déduites par la suite à l'aide du positionnement des plans de mise au point RVB.

Les deux détecteurs Dalsa et Stingray ont des dimensions différentes. En utilisant le même objectif successivement sur les deux détecteurs, il ne sera donc pas possible d'avoir le même champ d'observation. Nous choisissons donc une focale permettant de respecter l'intervalle fixé dans le cahier des charges à savoir un champ compris entre 20° et 30°. Comme le détecteur Dalsa est le plus grand, nous imposons d'avoir un champ de 30° avec ce détecteur. La distance focale est ensuite déterminée en utilisant la relation suivante :

$$f = \frac{\sqrt{N_{px,h}^2 + N_{px,v}^2}}{2\tan(\theta/2)} p_x, \quad (4.2)$$

où θ correspond au champ en diagonale. L'application numérique associée à un champ d'observation en diagonale de 30° donne une distance focale de 24 mm. Avec le détecteur Stingray ceci conduit à un champ en diagonal de 26°, ce qui respecte également le cahier des charges. Notons que cette valeur de distance focale est proche des 25 mm des objectifs du commerce.

4.2.3.3 Ouverture

Notre cahier des charges n'impose pas de contrainte particulière sur le nombre d'ouverture, si ce n'est que le drone se déplace dans un environnement urbain qui peut être faiblement éclairé. Or un nombre d'ouverture $F/\#$ trop élevé réduit le flux incident sur le détecteur, ce qui dégrade le rapport signal à bruit de l'image notamment lorsque l'illumination de la scène est faible. En contrepartie, une valeur de $F/\#$ trop faible augmente les aberrations, ce qui complexifie la conception optique. Nous fixons donc cette ouverture à 3, car cette valeur nous semble un bon compromis pour l'acquisition de scènes en intérieur ou en extérieur plus ou moins bien éclairées, et correspond à l'ordre de grandeur de l'ouverture utilisée pour les validations expérimentales réalisées dans les chapitres précédents.

4.2.3.4 Taille de la fenêtre de données traitée par l'algorithme

Dans le cahier des charges la résolution de la carte de profondeur est fixée autour de 2 cm à 3 m. Or comme nous l'avons mentionné à la section 4.1.2, la résolution de la carte de profondeur influence la taille de la fenêtre de donnée traitées par l'algorithme. En effet, pour une fenêtre de données de taille $T_{max} \times T_{max}$ pixels, la résolution de la carte de profondeur est donnée par la relation :

$$\delta_x = \frac{T_{max} p_x}{f} p. \quad (4.3)$$

Pour une résolution de la carte de profondeur fixée autour de 2 cm à 3 m et une distance focale de 24 mm, la taille maximale de la fenêtre traitée doit être de 21 pixels pour le détecteur Dalsa et de 45 pixels pour le détecteur Stingray. Les cartes de profondeur présentent alors 48×66 points de mesures pour le détecteur Dalsa et 45×54 points de mesures pour le détecteur Stingray, sans recouvrement des fenêtres traitées par l'algorithme.

4.2.3.5 Bornes sur la position des plans de mise au point RVB

En DFD, une règle empirique impose que la FEP soit de taille inférieure à la taille de la fenêtre de données traitées par l'algorithme. Ceci implique donc une contrainte sur la position maximale du plan de mise au point du canal rouge et la position minimale du plan de mise au point du canal bleu. En effet, comme illustré par exemple dans les figures 4.2(a) et (b), pour une position donnée des plans de mise au point RVB d'un imageur chromatique,

les FEP les plus larges correspondent respectivement à la FEP du canal rouge, pour une faible profondeur, et à la FEP du canal bleu, pour des grandes profondeurs. Ces FEP sont d'autant plus larges que le plan de mise au point rouge est éloigné de l'imageur et que le plan bleu est proche. Connaissant la taille maximale de la fenêtre de données, donc la largeur maximale de la FEP ϵ_{\max} , et d'après la formule (1) relative au diamètre géométrique de la FEP, présentée à la page 4 de l'introduction, nous pouvons donc déduire les positions extrêmes des plans de mise au point des canaux rouge et bleu par les relations :

$$\frac{1}{p_{0R,\max}} \simeq \frac{1}{p_{\min}} - \frac{\epsilon_{\max} F_{/\#}}{f^2} \quad (4.4)$$

$$\frac{1}{p_{0B,\min}} \simeq \frac{1}{p_{\max}} + \frac{\epsilon_{\max} F_{/\#}}{f^2} \quad (4.5)$$

$$(4.6)$$

où $F_{/\#}$ est le nombre d'ouverture. Pour une résolution de la carte de profondeur fixée autour de 2 cm à 3 m et une distance focale de 24 mm, la taille maximale de la FEP ϵ_{\max} est de l'ordre de 160 μm . Avec un nombre d'ouverture $F_{/\#}$ fixé à 3, nous obtenons alors des plans de mise au point des canaux R et B respectivement inférieur à 6 m et supérieur à 1 m, ce qui laisse une assez grande liberté sur le positionnement des plans de mise au point.

4.2.4 Dimensionnement préliminaire des plans de mise au point RVB

Pour dimensionner la position des plans de mise au point RVB, nous comparons les performances de plusieurs systèmes optiques simulés à partir des paramètres que nous avons fixés dans la section 4.2.3, à savoir la distance focale, l'ouverture, la fenêtre de données traitées par l'algorithme et les bornes des plans de mise au point des canaux R et B. Ce prédimensionnement est réalisé avec une modélisation très simple des FEP par des fonctions gaussiennes. Nous présentons les résultats de ce prédimensionnement pour le détecteur Dalsa puis pour le détecteur Stingray.

4.2.4.1 Liste de triplets de mise au points des canaux RVB potentiels

La première étape du dimensionnement préliminaire consiste à construire un ensemble de triplets de mise au point potentiels pour l'imageur chromatique, compte tenu des bornes définies à la section 4.2.3.5. La position maximale du plan de mise au point rouge est alors fixée à 6 m et la position minimale du plan de mise au point du canal bleu est fixée à 1 m. Nous générons ensuite un ensemble de triplets de mise au point potentiels $[p_0^B \ p_0^V \ p_0^R]$ avec $p_0^B \leq p_0^V \leq p_0^R$ ¹. Pour limiter le nombre de triplets potentiels nous séparons les trois plans RVB d'au minimum 20 cm. Nous obtenons ainsi près de 4300 triplets potentiels dont nous comparons la performance en fonction des critères définis dans la section 4.2.2.

4.2.4.2 Modélisation de l'imageur

À partir d'un triplet de mise au point potentiel $[p_0^B \ p_0^V \ p_0^R]$, la distance focale de 24 mm est attribuée au canal vert ce qui permet de définir la position du détecteur par rapport à l'optique, soit la distance d_{\det} , en utilisant les formules d'optique géométrique des lentilles minces :

$$\frac{1}{p_0^C} + \frac{1}{d_{\det}} = \frac{1}{f_C}, \quad (4.7)$$

¹Nous supposons ici que le chromatisme est issu de l'effet de réfraction dans les lentilles. Si celui-ci était produit par des lentilles diffractives l'ordre des plans de mise au point RVB devrait être inversé.

où C correspond au canal R,V ou B. Puis, connaissant les paramètres d_{det} , p_0^B et p_0^R , les valeurs de distances focales f_R et f_B sont calculées pour les canaux R et B en utilisant l'équation (4.7). Ces distances focales permettent d'évaluer le chromatisme longitudinal, défini comme la différence entre f_R et f_B .

Connaissant la position du détecteur, la valeur du nombre d'ouverture, la taille des pixels et les distances focales des canaux RVB, il est alors possible de simuler les FEP de l'imageur. Pour des raisons de rapidité de calcul, nous choisissons de modéliser les FEP par des fonctions gaussiennes paramétrées par l'écart type τ tel que :

$$\tau = 2\rho \frac{Rd_{det}}{p_x} \left(\frac{1}{f} - \frac{1}{d_{det}} - \frac{1}{p} \right). \quad (4.8)$$

où ρ est un paramètre correctif qui dépend de l'imageur. Dans la littérature la valeur de ρ est souvent déterminée par calibrage [Rajagopalan et Chaudhuri, 1998]. Afin de fixer cette valeur de sorte que la taille de la FEP gaussienne soit cohérente avec la taille effective de la FEP, nous cherchons la valeur du paramètre ρ qui permet d'approcher au mieux la taille des FEP réelles. Pour cela, nous comparons les FEP de défocalisation simulées et les FEP gaussiennes pour différentes valeurs de ρ ce qui a aboutit à un choix arbitraire de ρ fixé à 0.31. La figure 4.3 présente une coupe horizontale de deux FEP simulées à 3 m pour une mise au point à 1.6 m et $\rho = 0.31$, en utilisant une fonction gaussienne ou en utilisant les formules d'optique physique. Cette figure illustre que le choix de ρ permet d'avoir une tache de défocalisation gaussienne de largeur proche de la tache de défocalisation simulée par les formules d'optique de Fourier.

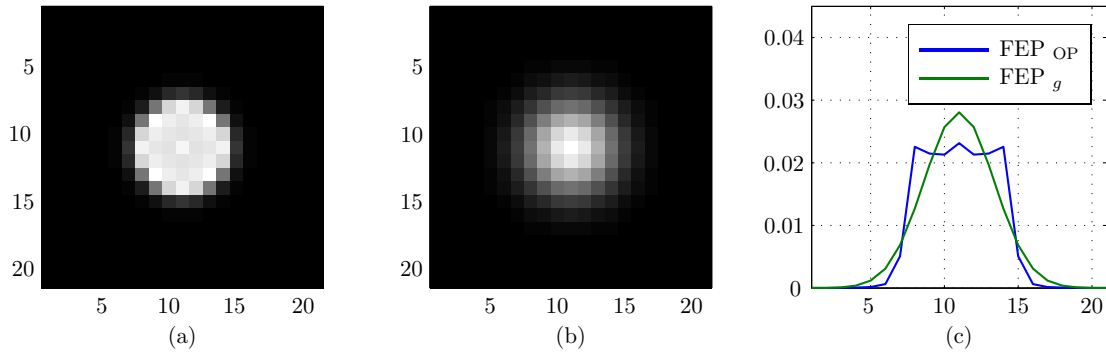


Fig. 4.3: Comparaison des allures des FEP simulées : (a) par les formules d'optique physique (FEP_{OP}), (b) par une fonction gaussienne (FEP_g). Le plan de mise au point est situé à 1.6 m, $\rho=0.31$ et les FEP sont simulées pour une profondeur de 3 m avec les paramètres du détecteur Dalsa et les paramètres de l'imageur chromatique fixés à la section 4.2.3. (c) Coupe horizontale des FEP pour les deux modélisations.

4.2.4.3 Comparaison exhaustive des triplets de mise au point potentiels

Dans un premier temps nous calculons les critères C1 et C2, pour chacun des triplets potentiels construits à la section 4.2.4.1. Rappelons que le critère C1 caractérise la précision moyenne sur la portée d'estimation de profondeur de l'imageur et que C2 caractérise la profondeur de champ de l'image restaurée. Pour cela, pour chaque triplet potentiel, nous simulons les FEP associées aux profondeurs comprises entre 1 et 5 m et en déduisons les valeurs de σ_{BCR} . Ceci permet de calculer la valeur du critère C1. Notons que nous utilisons

dans un premier temps une valeur du paramètre de régularisation α , défini comme l'inverse du rapport signal à bruit à la section 1.2.5 page 27, égale à 0.001 mais nous montrons ensuite que l'ordre de grandeur des plans de mise au point optimaux varie peu pour d'autres valeurs de ce paramètre. Ensuite, pour chaque triplet, compte tenu des paramètres optiques tels que la distance focale, l'ouverture et la taille des pixels, nous calculons le critère C2 en utilisant les formules caractérisant la taille géométrique des FEP des canaux RVB en fonction de la profondeur, donnés par l'équation (1) page 4.

Après avoir calculé les deux critères pour chaque triplet potentiels, nous déterminons alors le triplet de mise au point minimisant le critère C1, donc la valeur moyenne de σ_{BCR} , ainsi que le triplet maximisant le critère C2. Il s'est avéré que triplets optimaux suivant chacun des deux critères différent, il est donc nécessaire de déterminer un triplet correspondant à un compromis entre ces deux critères. Pour cela, nous sélectionnons l'ensemble des triplets dont la valeur du critère C1 correspond à une variation de moins de 5% de la valeur de C1 la plus faible et choisissons parmi ces triplets celui qui a la plus haute valeur du critère C2. Ce système est alors considéré comme un compromis entre les deux critères C1 et C2.

4.2.4.4 Dimensionnement préliminaire pour le détecteur Dalsa

Le tableau 4.2 présente respectivement le triplet de mise au point minimisant le critère C1, celui maximisant le critère C2 et le triplet correspondant à un compromis entre ces deux critères. Les figures 4.4, 4.5 et 4.6 présentent les courbes de taille de FEP théoriques et de précision d'estimation de profondeur pour les trois triplets sélectionnés.

Critère	$p_0^B(\text{m})$	$p_0^V(\text{m})$	$p_0^R(\text{m})$	C1 (mm)	C2 (m)
C1	2.4	3.4	6.0	7.4	3
C2	1.8	2.4	3.8	13.2	3.4
Compromis	2.2	3.2	6	7.5	3.1

Tableau 4.2: Triplets de mise au point maximisant la précision d'estimation de profondeur (critère C1), maximisant la capacité à produire une image de grande profondeur de champ (critère C2) et triplet présentant un compromis entre les deux critères.

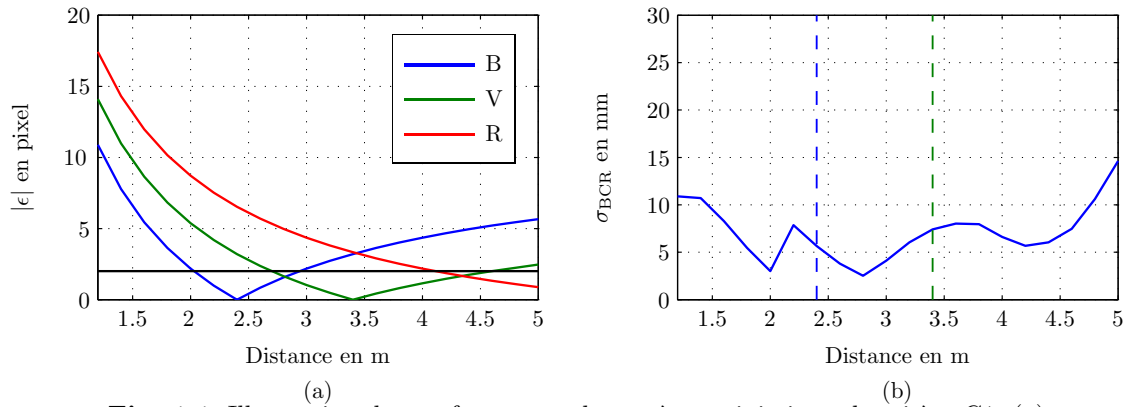


Fig. 4.4: Illustration des performances du système minimisant le critère C1. (a) Variation de la taille géométrique des FEP pour les trois canaux RVB en fonction de la profondeur. (b) Variation du critère σ_{BCR} en fonction de la profondeur.

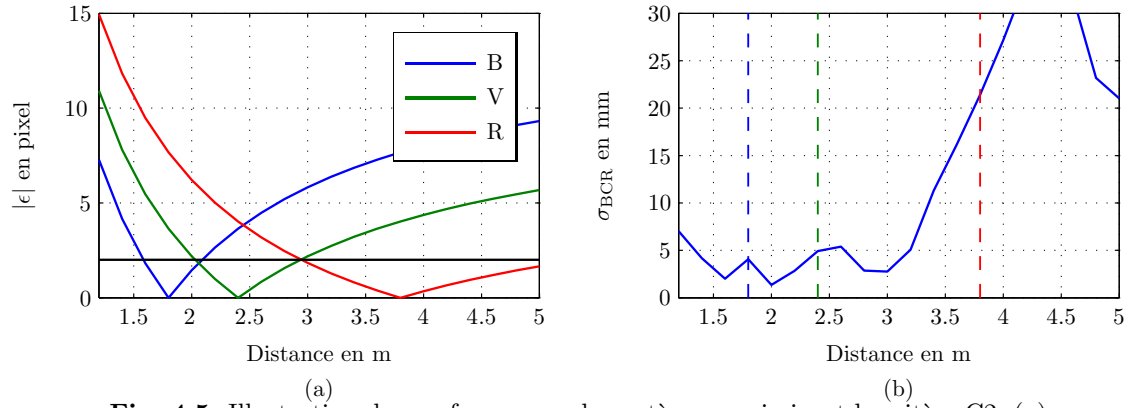


Fig. 4.5: Illustration des performances du système maximisant le critère C2. (a) Variation de la taille géométrique des FEP pour les trois canaux RVB en fonction de la profondeur. (b) Variation du critère σ_{BCR} en fonction de la profondeur.

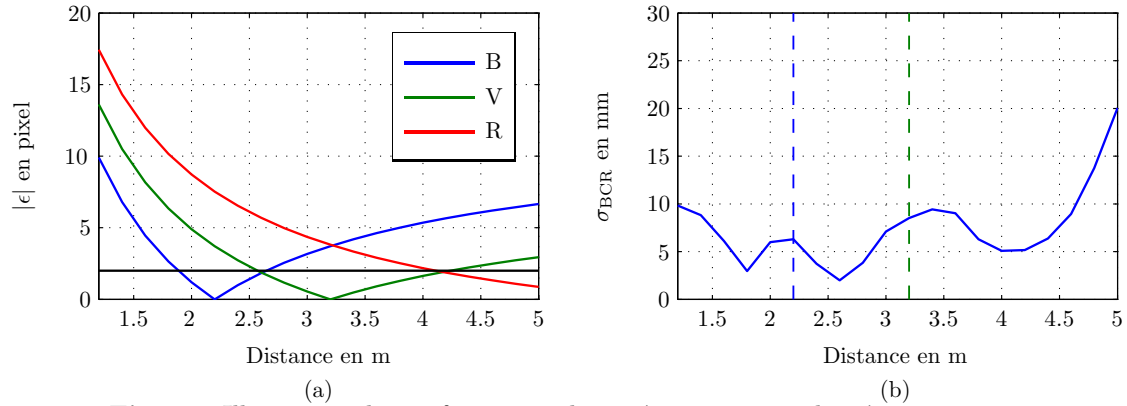


Fig. 4.6: Illustration des performances du système correspondant à un compromis entre les critères C1 et C2. (a) Variation de la taille géométrique des FEP pour les trois canaux RVB en fonction de la profondeur. (b) Variation du critère σ_{BCR} en fonction de la profondeur.

La figure 4.4 présente les performances liées au triplet maximisant la précision, autrement dit minimisant le critère C1. La figure 4.4(a) montre qu'il existe toujours une FEP de largeur inférieure à 2 pixels entre environ 2 m et 5 m, ce qui montre qu'un transfert de hautes fréquences est possible dans cette région. Sur la figure 4.4(b) nous pouvons observer que les valeurs de σ_{BCR} sont toujours inférieures à 15 mm avec une augmentation locale au niveau des plans de mise au point des canaux RVB. La figure 4.5(a) illustre que pour le triplet maximisant le critère C2, les tailles de FEP théoriques se croisent au niveau de la limite fixée à 2 pixels, ce qui permet d'avoir toujours au moins une des FEP inférieure à 2 pixels, donc un transfert de fréquences possible, entre 1.7 m et 5 m. En contrepartie, nous pouvons observer à la figure 4.5(b) une forte augmentation des valeurs de σ_{BCR} notamment pour des profondeurs comprises entre 3.5 m et 5 m. Ceci peut s'expliquer par le fait que les FEP des canaux B et V sont alors plus larges que pour le triplet maximisant la précision et nous avons vu dans le chapitre 3 que la précision d'estimation de profondeur est plus importante pour des petites FEP. La figure 4.6 présente les performances du triplet que nous considérons comme un compromis entre les critères C1 et C2. La figure 4.6(a) illustre l'intérêt de cette configuration vis-à-vis des deux triplets précédents. En effet, au moins l'une des FEP est inférieure à 2 pixels dans la région comprise entre 1.8 m et 5 m et par rapport au triplet maximisant le critère C2, le pic des valeurs de σ_{BCR} apparaissant pour

les profondeurs importante est atténué.

En conclusion, cette étude montre que les triplet qui minimisent le critère C1 et maximisent le critère C2 sont différents. La minimisation du critère C1 favorise les FEP de tailles réduites pour les trois canaux dans une large plage de profondeur. Par contre, maximiser le critère C2 tend à avoir des plans de mise au point répartis dans la plage de distance afin d'avoir un croisement des courbes de tailles de FEP autours de 2 pixels, quitte à avoir de larges FEP et donc une dégradation de la précision. Le triplet correspondant à un compromis permet d'améliorer un peu le critère C2 sans trop dégrader la précision d'estimation de profondeur.

L'étude précédente a été réalisée pour une valeur du paramètre de régularisation α , autrement dit pour une valeur de rapport signal à bruit. Les simulations précédentes sont donc reproduites pour différentes valeurs de ce paramètre. Les triplets de mise au point correspondant à un compromis entre les critères C1 et C2 obtenus sont présentés dans le tableau 4.3.

α	$p_0^B(\text{m})$	$p_0^V(\text{m})$	$p_0^R(\text{m})$	C1 (mm)	C2 (m)
0.1	2.2	3.4	4.8	22.5	3.1
0.01	2.2	3.4	6.0	12.4	3.1
0.001	2.2	3.2	6.0	7.5	3.1
0.0001	2.2	3.2	6.0	5.4	3.1

Tableau 4.3: Triplets correspondant à un compromis entre les critères C1 et C2 pour différentes valeurs du paramètre α .

Les plans de mise au point correspondant à un compromis entre les critères C1 et C2 varient peu sauf pour un bruit très important. Ces résultats montrent donc une certaine robustesse de la position des plans de mise au point du système présentant un compromis entre qualité de l'image et précision, vis-à-vis du rapport signal à bruit.

4.2.4.5 Dimensionnement préliminaire pour le détecteur Stingray

Nous reproduisons l'étude précédente pour le détecteur Stingray. Le tableau 4.4 présente les plans de mise au point optimaux vis-à-vis des critères C1 et C2 et le triplet correspondant à un compromis entre ces deux critères.

Critère	$p_0^B(\text{m})$	$p_0^V(\text{m})$	$p_0^R(\text{m})$	C1 (mm)	C2 (m)
C1	2.2	3	5.6	3.7	1.3
C2	2.6	3.2	4.2	4.7	2.5
Compromis	2.4	3	5.6	4.1	2.2

Tableau 4.4: Plans de mise au point optimaux suivant les critères C1, C2 et un compromis entre les deux critères pour le détecteur Stingray.

Les résultats du tableau 4.4 montrent que les plans de mise au point optimaux, en termes de précision, sont assez voisins de ceux obtenus avec le détecteur Dalsa. En revanche le triplet de mise au point optimal vis-à-vis du critère C2 a changé et la valeur du critère pour le triplet optimal a diminué. Ce résultat n'est pas surprenant car en diminuant la taille des pixels nous avons augmenté la taille des FEP donc la plage de distance pour laquelle la FEP est inférieure à deux pixels a diminué. La position maximisant le critère C2 est donc

différente et la valeur du critère diminue. Enfin, le triplet correspondant à un compromis présente des plans de mise au point proches de ceux obtenus pour le détecteur Dalsa. Cette étude indique que la position des plans de mise au point optimaux en termes de précision est peu sensible à la taille des pixels du détecteur.

4.2.4.6 Conclusion

L'objectif de cette étude était de déterminer un ordre de grandeur de la position des plans de mise au point RVB de l'objectif chromatique co-conçu. Cette étude montre que pour les deux détecteurs, des plans de mise au point des canaux R, V et B situés respectivement entre 5 et 6 m, 3.2 m et 2.2 m permettent d'avoir un bon compromis entre la précision d'estimation de profondeur et la reconstruction d'une image de grande profondeur de champ. En associant la distance focale du canal vert à 24 mm et en utilisant la formule (4.7), nous pouvons en déduire que ces plans correspondent à un chromatisme longitudinal de l'ordre de 150 μm .

Nous avons pu constater lors du prédimensionnement que la position des plans de mise au point correspondant à un compromis entre la précision et la restauration d'une image RVB de grande profondeur de champ est peu sensible à la taille du détecteur, en tout cas pour les caractéristiques des deux détecteurs testés lors de l'étude. Dans la suite de ce chapitre, nous utilisons donc les caractéristiques du détecteur Dalsa, en faisant l'hypothèse que l'optimisation de l'objectif chromatique sera également favorable lorsque ce dernier est utilisé avec le détecteur Stingray.

4.2.5 Conception d'une première architecture optique

À partir du premier dimensionnement réalisé sur un imageur idéal simulé, une première architecture d'imageur chromatique a été conçue, en collaboration avec le DOTA, à l'aide du logiciel de conception optique **Zemax**. Ce logiciel permet d'optimiser les paramètres des lentilles constituant un objectif, à savoir le rayon de courbure, l'épaisseur, la position et le type de verre, en fonction de critères reposant sur la qualité image.

4.2.5.1 Choix d'une architecture optique

Pour concevoir l'objectif chromatique co-conçu, nous avons choisi de contrôler le chromatisme grâce à l'utilisation de lentilles d'indice de réfraction et de constringence différents. Notons qu'un autre moyen d'agir sur le chromatisme aurait consisté à utiliser une lentille diffractive mais cette solution a été écartée à cause de la variabilité de l'efficacité de diffraction avec la longueur d'onde. Nous souhaitons en effet faire de l'imagerie sur une large de bande spectrale entre 400 et 700 nm. De plus, nous avons choisi de rendre la pupille accessible, afin de pouvoir envisager l'introduction d'une pupille codée a posteriori dans l'imageur chromatique. Afin de limiter l'aberration chromatique latérale et la distorsion, les lentilles sont réparties de part et d'autre de la pupille. Enfin, l'imageur est conçu pour être télécentrique image. Cette propriété consiste à positionner la pupille de sortie de l'imageur à l'infini. Ainsi, le barycentre de la FEP associée à un objet atteint le détecteur à la même position quelle que soit la position du détecteur. Ceci permet de pouvoir modifier la position du détecteur sans modifier le grandissement. Pour limiter l'encombrement et la complexité de l'imageur, le nombre de lentilles est limité à 4.

4.2.5.2 Définition d'une fonction de mérite

Le logiciel de conception optique **Zemax** permet de définir une fonction de mérite caractérisant la qualité image, fonction qui est ensuite utilisée pour l'optimisation des paramètres

des lentilles. Pour l'optimisation de l'objectif chromatique, nous avons ajouté à la fonction de mérite initiale de **Zemax** des contraintes visant à corriger plusieurs aberrations optiques. Tout d'abord nous minimisons l'aberration sphérique car cette aberration diminue la variabilité de la FEP autour du plan de mise au point. Nous imposons également une contrainte permettant de maintenir une atténuation de l'illumination dans l'image inférieure à 10%. En effet, une diminution de l'intensité lumineuse dans le champ peut perturber l'algorithme d'estimation de profondeur. Par ailleurs, nous ajoutons dans la fonction de mérite le paramètre AXCL pour *Axial color*. En effet, ce paramètre caractérise l'écart entre les images paraxiales pour deux longueurs d'onde. En choisissant ces deux longueurs d'onde comme les longueurs d'onde centrales des spectres rouge et bleu des filtres de la Dalsa représentés à la figure 4.7, ce paramètre caractérise alors le chromatisme de l'imageur.

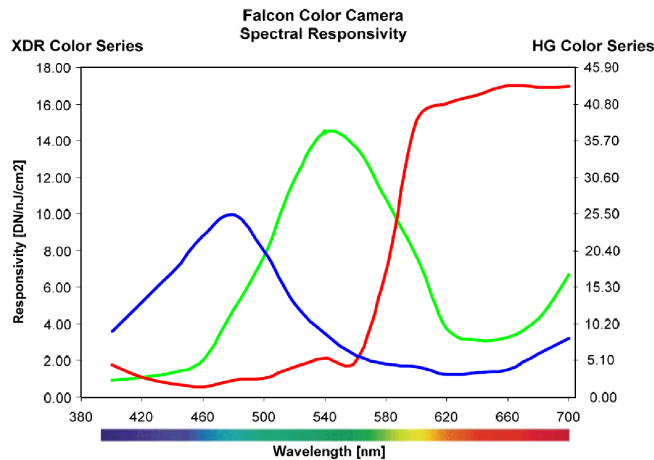


Fig. 4.7: Spectres des filtres RVB présents sur le détecteur Dalsa, d'après les données du constructeur.

4.2.5.3 Optimisation d'une première architecture

Après avoir défini la fonction de mérite, le logiciel **Zemax** optimise les paramètres des lentilles afin de minimiser cette fonction. Nous effectuons cette optimisation pour une ouverture de 2.5, afin de pouvoir envisager l'ajout d'une pupille codée de diamètre plus large que celui de la pupille circulaire correspondant à un nombre d'ouverture égal à 3, tout en maintenant constant le flux lumineux atteignant les pixels du détecteur. Nous utilisons alors un mode multi-configurations qui permet d'optimiser simultanément la qualité image des trois canaux de l'imageur. Cette première version de l'imageur chromatique est optimisée pour une valeur de AXCL fixée autour de 150 μm . La figure 4.8 présente l'architecture obtenue, constituée de quatre lentilles. Ce type d'architecture est voisin de l'architecture classique double Gauss, sauf que les doublets correspondent dans notre cas à une seule lentille. Afin d'obtenir une qualité image acceptable vis-à-vis de la fréquence de Nyquist du détecteur Dalsa, la surface 7 de cette architecture est asphérisée.

4.2.6 Optimisation conjointe

À partir de la première version de l'imageur, nous avons optimisé conjointement les paramètres de chaque surface, en utilisant les critères de performance définis à la section 4.2.2 portant sur la précision et sur la capacité à produire une image de grande profondeur de champ.

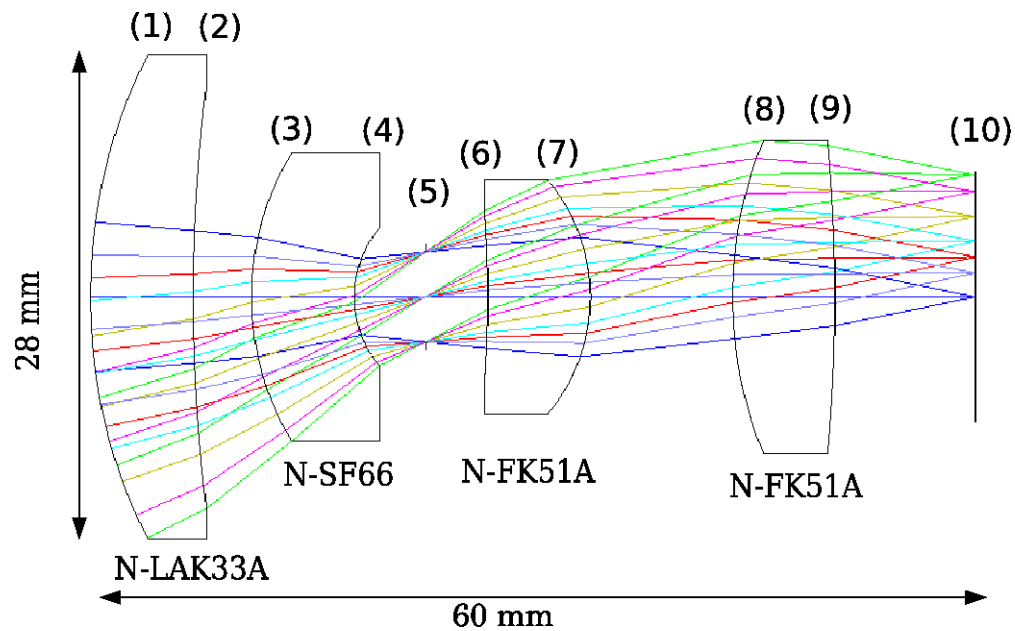


Fig. 4.8: Plan des lentilles de l'architecture choisie pour CAM3D sous Zemax. Les surfaces sont numérotées de 1 à 10. La surface (5) correspond à la pupille et la surface (10) au plan du détecteur. Nous indiquons pour chaque lentille la nature du verre.

4.2.6.1 Récupération des FEP simulées par Zemax

Afin d'évaluer la précision d'estimation de profondeur de l'imageur, il est nécessaire de récupérer les FEP simulées par Zemax pour différentes profondeurs. À l'aide de la librairie Matlab `zdde`, qui permet de lancer des fonctions de Zemax à partir de Matlab, les FEP polychromatiques des trois canaux RVB sont simulées sous Zemax pour différentes profondeurs et récupérées sous Matlab. Ceci permet alors de calculer la précision d'estimation de profondeur d'un imageur donné et donc le critère C1. D'autres paramètres sont également extraits à l'aide de cette librairie comme par exemple les valeurs de FTM à la fréquence de Nyquist autrement dit à la moitié de la fréquence d'échantillonnage. Ceci nous permet de tracer les courbes de variation de la FTM des trois canaux à la fréquence de Nyquist en fonction de la profondeur, comme illustré à la figure 4.10. Nous définissons alors la position des plans de mise au point RVB au niveau des maxima de ces courbes. La connaissance de la position de ces plans nous permet alors de calculer le critère C2.

4.2.6.2 Optimisation manuelle de l'objectif chromatique

L'insertion d'un critère de performance calculé sous Matlab tels que la BCR dans la fonction de mérite de Zemax est possible en théorie mais cela demande un certain travail de programmation qui n'a pas pu être réalisé dans les délais que nous avons fixés pour concevoir l'imageur chromatique. C'est pourquoi nous utilisons une optimisation manuelle dans laquelle à partir de l'architecture préliminaire, nous fixons une nouvelle valeur du paramètre AXCL, puis réoptimisons l'architecture. Pour chaque valeur de AXCL, nous caractérisons la performance de l'architecture obtenue pour différentes positions du détecteur, ce qui déplace les plans de mise au point RVB. Nous sélectionnons alors les triplets de plan de mise au point RVB minimisant le critère de précision C1 ou maximisant le critère de profondeur

de champ généralisé C2. Le tableau 4.5 présente pour chaque valeur du paramètre AXCL les plans de mise au point optimaux selon ces deux critères.

AXCL	critère	$p_0^B(\text{m})$	$p_0^V(\text{m})$	$p_0^R(\text{m})$	C1 en mm	C2 en m
50 μm	C1	2.1	2.3	2.7	5.0	1.5
	C2	2.8	3.1	3.8	6.4	2.7
85 μm	C1	2.2	2.55	3.3	5.4	2.4
	C2	2.4	2.8	3.7	5.8	3.0
100 μm	C1	2	2.4	3.2	5.9	2.3
	C2	2.2	2.7	3.7	6.4	3.1
121 μm	C1	3.0	2.45	3.35	6.4	2.6
	C2	2.1	2.65	3.8	6.8	3.2
130 μm	C1	1.8	2.25	3.1	6.5	2.3
	C2	2.1	2.6	3.8	7.2	3.2
157 μm	C1	1.7	2.2	3.1	7.3	2.5
	C2	1.9	2.55	3.9	7.9	3.3
180 μm	C1	1.8	2.6	4.3	8.5	3.1
	C2	1.7	2.4	3.8	10.0	3.3

Tableau 4.5: Triplets optimisant les critères C1 et C2 pour différentes valeurs du paramètre AXCL.

Comme lors du prédimensionnement, nous pouvons observer que les triplets de plans de mise au point permettant la meilleure estimation de profondeur sont différents de ceux permettant de restaurer une image nette sur une large plage de profondeurs.

À partir de ces résultats, nous avons choisi un compromis entre la profondeur de champ généralisée et la précision d'estimation de profondeur en choisissant une valeur de AXCL de 100 μm et des plans de mise au point RVB respectivement à 3.7 m, 2.7 m et 2.2 m. Nous pouvons remarquer que la position des plans optimaux diffère quelque peu de celle obtenue lors du prédimensionnement en utilisant une modélisation des FEP par des gaussiennes, ce qui n'est pas surprenant car le logiciel **Zemax** permet de simuler les FEP de l'imageur de manière bien plus physique, en tenant compte notamment des aberrations.

4.3 Caractéristiques de l'objectif chromatique co-conçu

Dans cette section, nous présentons les performances de l'objectif chromatique co-conçu du point de vue de la qualité image et de la précision d'estimation de profondeur.

4.3.1 Paramètres des lentilles

La figure 4.9 présente les caractéristiques des lentilles de l'architecture choisie après l'optimisation manuelle. Les numéros 1 à 9 correspondent aux numéros des surfaces indiquées sur le plan des lentilles de la figure 4.8.

4.3.2 Position des plans de mise au point RVB

La figure 4.10 présente, pour les trois canaux RVB, les variations de la taille effective de la FEP, définie par le diamètre en nombre de pixels contenant 80% de l'énergie encerclée et la valeur de la FTM à la fréquence de Nyquist du détecteur en fonction de la profondeur.

4.3 Caractéristiques de l'objectif chromatique co-conçu

Surf	Type	Comment	Radius	Thickness	Glass	Semi-Diameter	Conic
OBJ	Standard		Infinity	3800.000		1027.477	0.000
1	Standard		25.502	7.000	N-LAK33A	13.642	0.000
2	Standard		74.904	4.019		11.975	0.000
3	Standard		13.288	7.026	N-SF66	8.139	0.000
4	Standard		5.245	4.851		3.887	0.000
STO	Standard		Infinity	4.241		2.552	0.000
6	Standard		-56.255	7.000	N-FK51A	4.886	0.000
7	Standard		-8.349	9.702		6.585	-0.399
8	Standard		19.291	7.000	N-FK51A	8.813	0.000
9	Standard		-76.449	9.545		8.499	0.000
IMA	Standard		Infinity	-		7.000	0.000

Fig. 4.9: Caractéristiques des lentilles de l'objectif chromatique co-conçu fixées dans Zemax.

La fréquence de Nyquist f_N vaut ici $1/4p_x$, à cause de la présence du filtre de Bayer sur le détecteur.

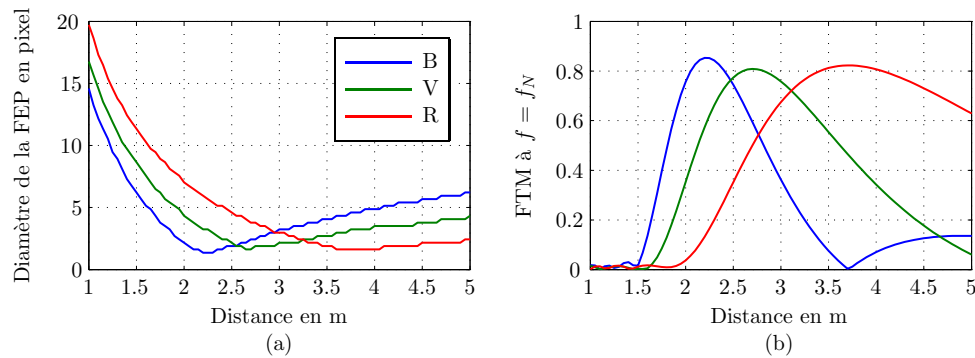


Fig. 4.10: (a) Diamètre contenant 80% de l'énergie encadrée de la FEP (b) Valeur de la FTM à la fréquence de Nyquist du détecteur Dalsa, soit 33 mm^{-1} , en fonction de la distance pour les trois canaux RVB.

Les plans de mise au point des canaux RVB de cet imageur, définis ici en repérant les profondeurs qui maximisent les courbes de FTM à 33 mm^{-1} , sont respectivement 2.2 m, 2.7 m et 3.7 m, ce qui est en accord avec les valeurs fixées lors de l'optimisation manuelle de l'imageur à la fin de la section 4.2.6.2.

4.3.3 Qualité de l'image hors axe

Les figures présentées dans cette section sont obtenues à partir du logiciel Zemax, pour l'objectif sélectionné par optimisation des critères de précision et de profondeur de champ à la section 4.2.6.2.

4.3.3.1 Variation de la FTM et de la FEP dans le champ

La figure 4.11 présente les variations des FTM en intensité des trois canaux RVB en fonction de la fréquence, pour un point source situé au niveau des plans de mise au point de chaque canal et pour plusieurs positions angulaires. Rappelons que pour le détecteur Dalsa, la fréquence de Nyquist est située à 33 mm^{-1} et qu'elle est située à 72 mm^{-1} pour le détecteur Stingray. La figure 4.11 montre que jusqu'à un champ de 10° les FTM des trois canaux sont supérieures à 0.5 à 33 mm^{-1} , ce qui permet d'avoir un bon contraste de l'image avec le détecteur Dalsa. Par contre, le détecteur Stingray aura un contraste moins bon car les FTM sont plus faibles à 72 mm^{-1} .

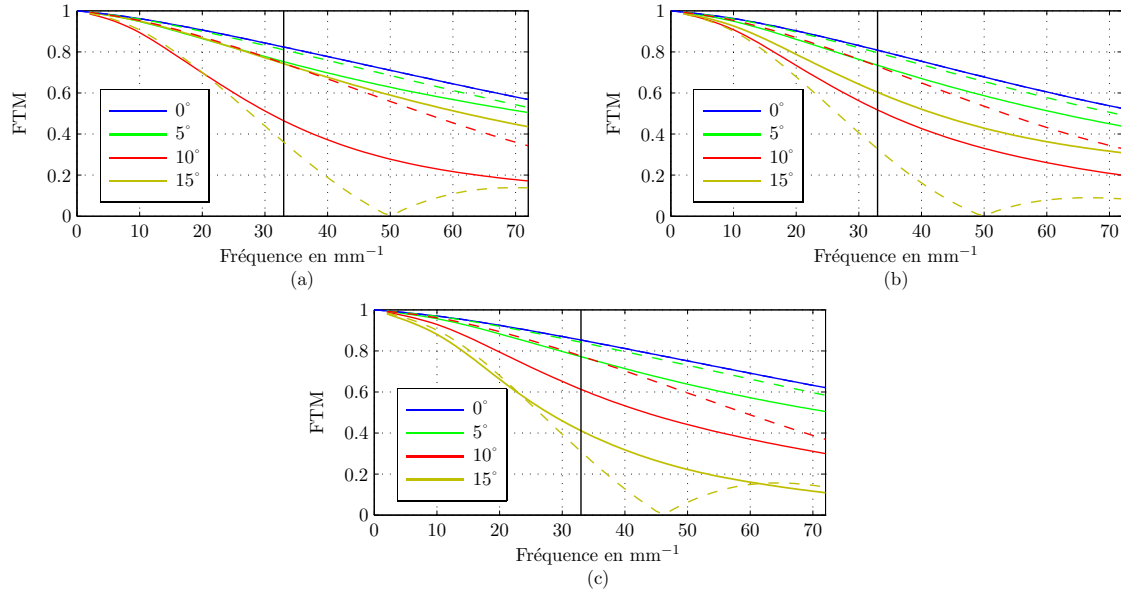


Fig. 4.11: (a) FTM du canal rouge à 3.7 m pour différentes valeurs du champ. (b) FTM du canal vert à 2.7 m. (c) FTM du canal bleu à 2.2 m. Le trait noir correspond à la fréquence de Nyquist du détecteur Dalsa, située à 33 mm^{-1} .

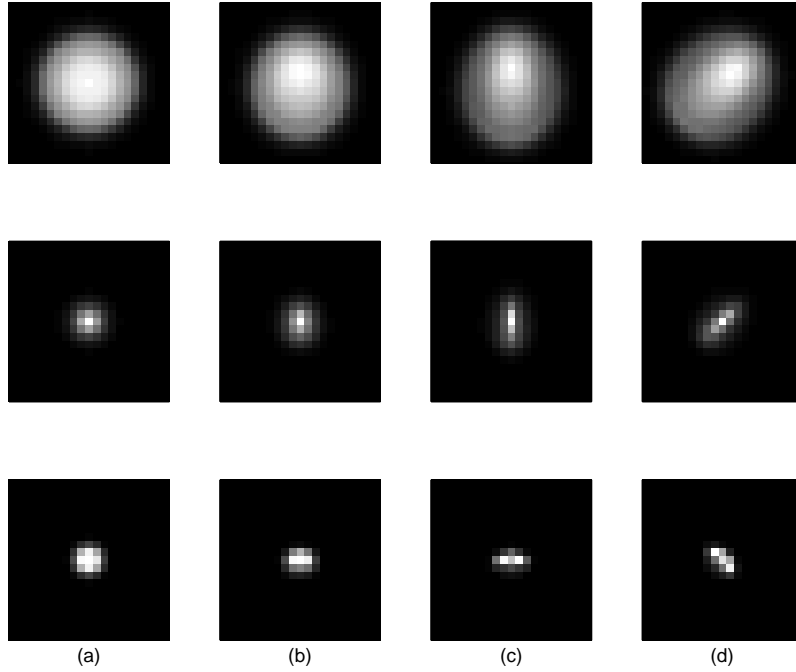


Fig. 4.12: (a) FEP à $(0^\circ, 0^\circ)$. (b) FEP à $(0^\circ, 5^\circ)$. (c) FEP à $(0^\circ, 10^\circ)$. (d) FEP à $(7^\circ, 7^\circ)$. De haut en bas les profondeurs sont respectivement 1.2 m, 1.8 m et 4.4 m. Le pas d'échantillonnage est de $7.4 \mu\text{m}$.

La figure 4.12 présente des exemples de FEP pour le canal vert à différentes profondeurs, sur l'axe et pour différentes valeurs de champ. Hors axe nous pouvons observer une déformation de la FEP dans une direction privilégiée. Ce comportement est typique d'une aberration d'astigmatisme qui est effectivement, d'après Zemax, une des aberrations princi-

pales de l'objectif. Cette déformation implique qu'il faudra envisager un calibrage des FEP hors axe pour l'estimation de profondeur sur une image entière.

4.3.3.2 Aberrations chromatique longitudinale et latérale

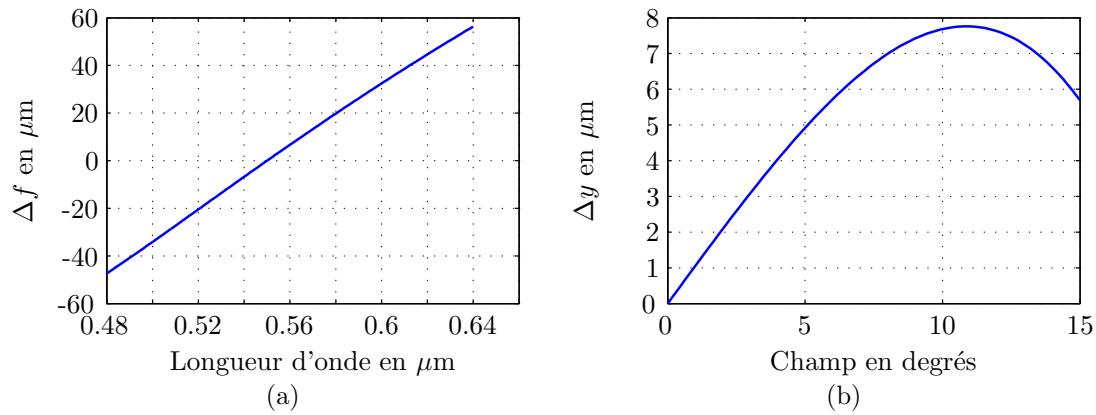


Fig. 4.13: (a) Variation de la distance focale Δf en fonction de la longueur d'onde, par rapport la référence du canal vert à 0.55 μm . (b) Variation de la position du point image Δy en fonction de l'angle de champ du point source.

La figure 4.13 (a) caractérise l'aberration chromatique longitudinale de notre objectif. Elle présente en effet la variation de l'écart de distance focale en fonction de la longueur d'onde par rapport à la distance focale du canal vert à 0.55 μm . Nous pouvons retrouver sur la figure (a) la valeur de chromatisme longitudinal de l'ordre de 100 μm .

Par ailleurs, la figure 4.13 (b) représente le chromatisme latéral de l'imageur en fonction de la longueur d'onde, autrement dit la variation de la position du centre de la FEP en fonction de l'angle de champ dans lequel se trouve le point source, en considérant qu'il est situé au niveau du plan de mise au point du canal vert. Cette variation reste inférieure à 8 μm , donc inférieure à un pixel du détecteur Dalsa et à environ 2 pixels du détecteur Stingray. Nous considérons cet écart comme suffisamment faible pour être négligé dans les traitements appliqués aux images produites par cet objectif avec ces deux détecteurs.

4.3.4 Performance d'estimation de profondeur sur l'axe

La figure 4.14 présente la performance d'estimation théorique de profondeur sur l'axe de l'objectif co-conçu avec les deux détecteurs Dalsa et Stingray. Nous pouvons remarquer que la précision d'estimation de profondeur théorique est meilleure avec le détecteur Stingray.

4.3.5 Réalisation de l'imageur

4.3.5.1 Tolérancement et réalisation des lentilles

Une étude du tolérancement est réalisée afin de déterminer les erreurs de fabrication de l'objectif chromatique qui peuvent être tolérées sans que la qualité optique ne soit trop dégradée. Nous avons notamment décidé d'accepter une chute de contraste maximale de 0.3 à la fréquence de Nyquist pour le détecteur Dalsa. Ce tolérancement porte sur les épaisseurs, les rayons de courbures des lentilles, le décentrement des différentes surfaces entre elles, la conicité et la position des lentilles. Cette étude a montré que la qualité optique sur l'axe et hors axe restait acceptable en tolérant une incertitude de $\pm 0.1\%$ sur les rayons de courbure

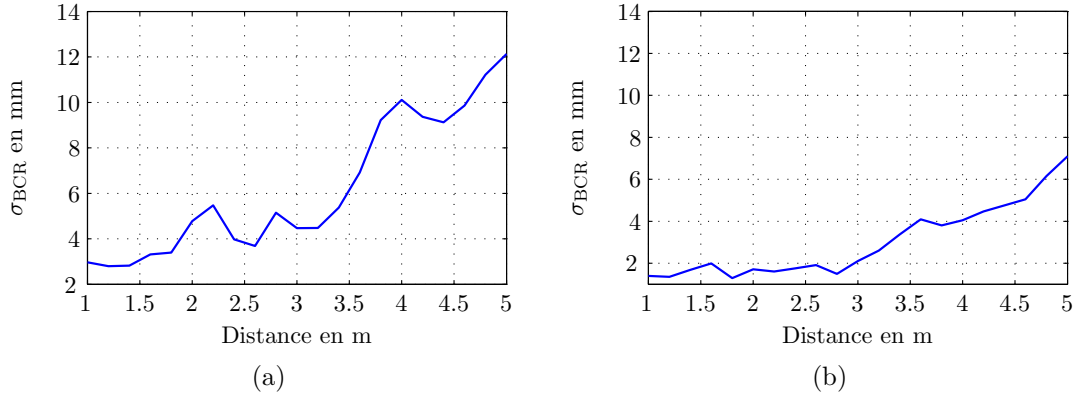


Fig. 4.14: Performance théorique d'estimation de profondeur de l'objectif co-conçu (a) avec le détecteur Dalsa (b) avec le détecteur Stingray.

et de $\pm 0.8\%$ sur la conicité de la surface arrière de la troisième lentille. La tolérance sur les erreurs d'épaisseur varie entre $50 \mu\text{m}$ à $100 \mu\text{m}$ et les erreurs tolérées sur le décentrement des surfaces sont de $20 \mu\text{m}$ et le décentrement des éléments varie de $25 \mu\text{m}$ à $50 \mu\text{m}$. Notons que les erreurs de décentrement sont les erreurs les plus critiques. À partir de ces contraintes, les lentilles non coniques ont été réalisées par la société R2B et la lentille conique par la société Eldim.

4.3.5.2 Mécanique

La distance entre le détecteur et la dernière lentille de l'objectif chromatique est trop faible pour pouvoir utiliser la monture initiale de la Dalsa ou de la Stingray qui sont des montures de type C. C'est pourquoi nous avons fait réaliser une monture dédiée à l'objectif chromatique. Cette monture a été conçue au DOTA et réalisée par la société Façon Mécanique avec une précision de l'ordre de $10 \mu\text{m}$. Les figures 4.15 et 4.16 présentent respectivement la monture conçue au DOTA et des photographies de l'objectif chromatique associé au détecteur Dalsa et au détecteur Stingray.

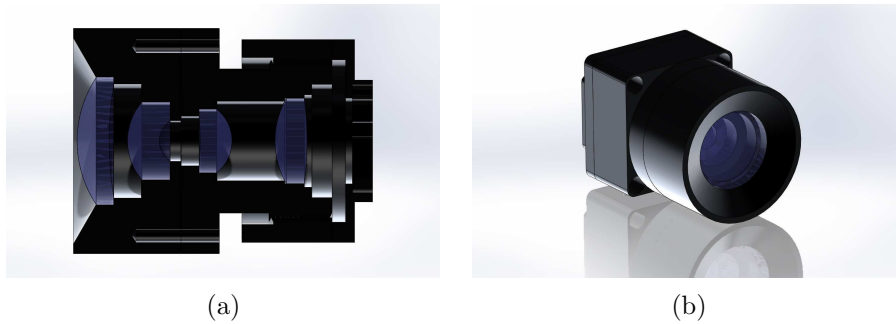


Fig. 4.15: Monture conçue au DOTA pour l'objectif chromatique.

4.4 Caractérisation de la précision expérimentale

Dans cette section, nous conduisons une étude de la précision d'estimation de profondeur expérimentale obtenue sur l'axe avec l'imageur chromatique co-conçu combinés aux deux détecteurs.



Fig. 4.16: Photographie de l'objectif chromatique co-conçu combiné : (a) au détecteur Dalsa (b) au détecteur Stingray.

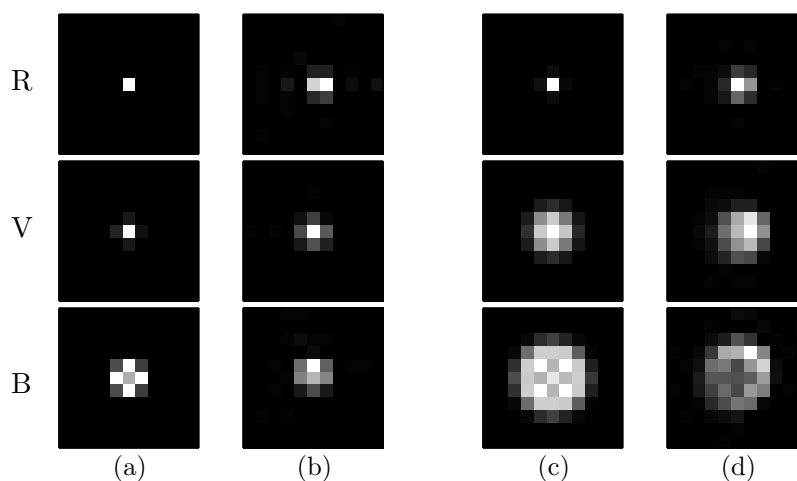


Fig. 4.17: Comparaison des FEP simulées à partir du logiciel *Zemax* et des FEP calibrées par la méthode de [Delbracio *et al.*, 2011], pour un point source situé à 4 m. (a) et (b) FEP simulées et calibrées avec le détecteur Dalsa. (c) et (d) FEP simulées et calibrées avec le détecteur Stingray. Les FEP sont celles produites sur chaque canal R, V et B.

4.4.1 Calibrage des FEP sur l'axe

Les FEP obtenues en combinant l'objectif co-conçu et les détecteurs Dalsa et Stingray sont calibrées sur l'axe en utilisant la méthode présentée dans [Delbracio *et al.*, 2011], pour la mise au point choisie lors de la conception, à savoir des plans de mise au point RVB situés respectivement à 3.7, 2.7 et 2.2 m. Dans chaque cas les FEP sont calibrées sur l'axe avec un pas de 5 cm. La figure 4.17 présente quelques exemples de FEP calibrées comparées aux FEP simulées sous *Zemax*. Nous pouvons d'abord observer que les tailles de FEP calibrées sont cohérentes avec les simulations sous *Zemax*. De plus, lorsque la FEP est de l'ordre de 1 pixel sous *Zemax*, elle s'étale sur plus de pixels sur les FEP réelles. Ceci peut provenir d'un écart entre la position théorique du détecteur et de la position réelle, ce qui décale les plans de mise au point, ou de la présence d'aberration sphérique introduite lors de la réalisation de l'imageur.

4.4.2 Précision sur l'axe avec le détecteur Dalsa

La figure 4.18 présente les résultats d'estimation de profondeur sur l'axe obtenus pour les cibles « Fleurs », « Ville », « Parterre » et « Carreaux » présentées à la figure 2.6 page 62. L'algorithme SAC-GLDFD traite une fenêtre de taille 21×21 pixels à l'échelle du détecteur. L'estimation de profondeur est réalisée sur une région centrée de taille 120×120 pixels avec un recouvrement de 50%.

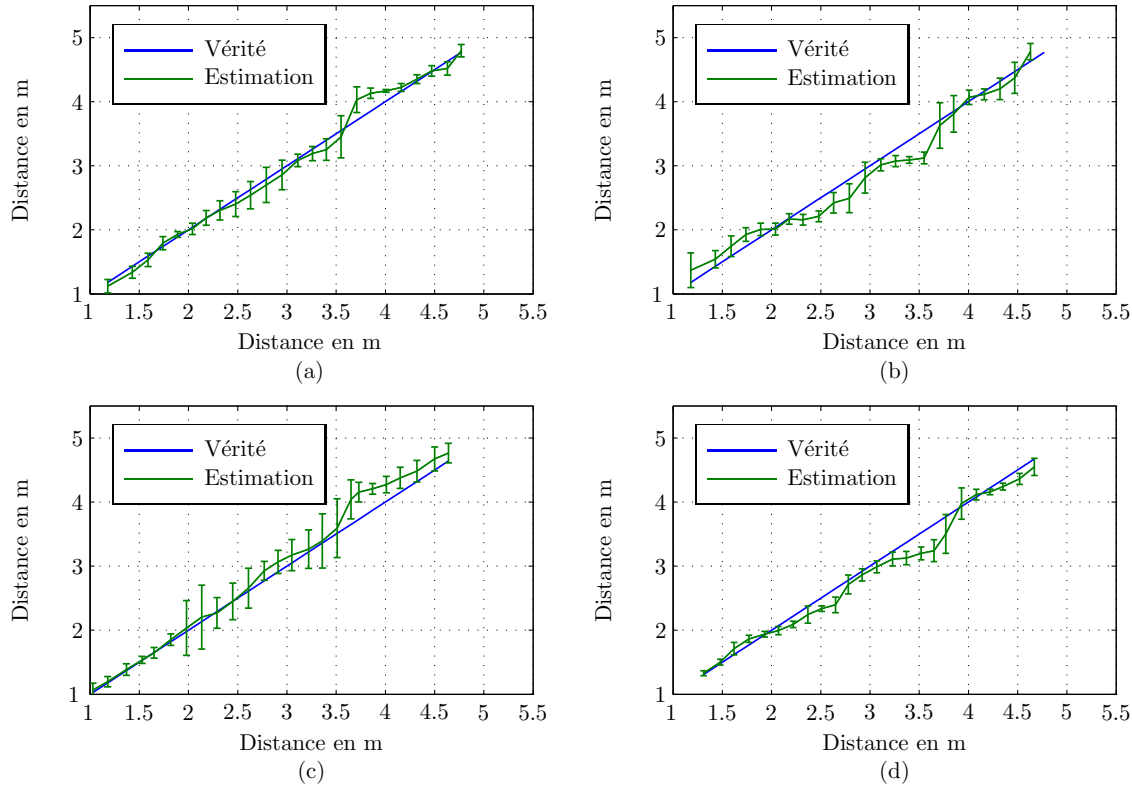


Fig. 4.18: Résultats d'estimation de profondeur obtenus sur l'axe avec l'objectif chromatique co-conçu et le détecteur Dalsa pour les cibles (a) « Fleurs », (b) « Ville », (c) « Parterre », (d) « Carreaux ».

Les résultats obtenus avec le détecteur Dalsa sont globalement cohérents avec la vérité donnée par le télémètre. Cependant, nous pouvons observer la présence d'un biais non négligeable, notamment pour les profondeurs comprises entre 2 m et 4 m.

4.4.3 Précision sur l'axe avec le détecteur Stingray

La figure 4.19 présente les résultats d'estimation de profondeur obtenus avec le détecteur Stingray. L'algorithme SAC-GLDFD traite une fenêtre de taille 45×45 pixels à l'échelle du détecteur. L'estimation de profondeur est réalisée sur une région centrée de taille 240×240 pixels avec un recouvrement de 50%. Le tableau 4.6 présente l'écart-type moyen et le biais moyen, pour chaque cible. Comme attendu après l'étude théorique présentée à la figure 4.14, nous obtenons avec le détecteur Stingray un écart-type expérimental plus faible qu'avec le détecteur Dalsa. Par ailleurs, comme le montre le tableau 4.6, l'écart-type moyen est de l'ordre de 7 cm sur l'ensemble des profondeurs considérées pour les quatre cibles et la valeur moyenne du biais est typiquement de 4 cm, ce qui est plus faible que lors de l'utilisation du détecteur Dalsa. Enfin la robustesse à la cible est meilleure avec le détecteur Stingray.

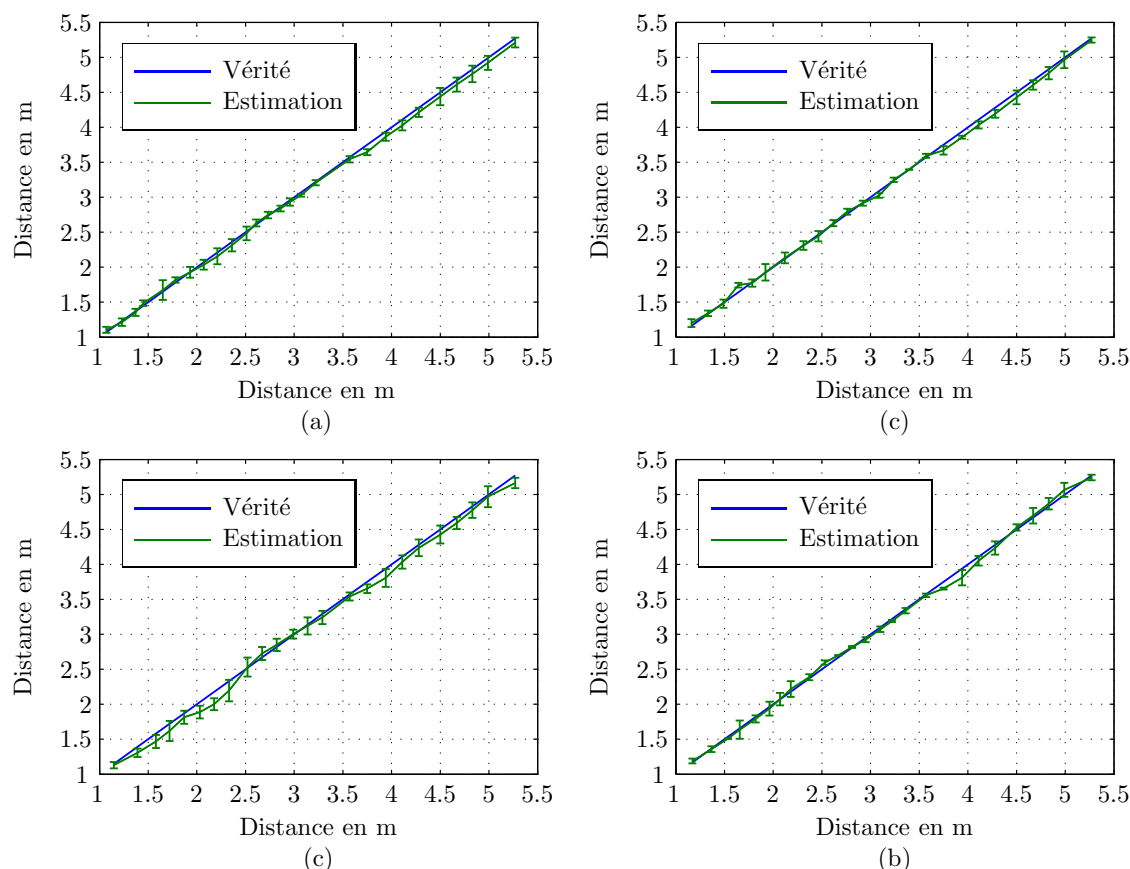


Fig. 4.19: Résultats d'estimation de profondeur obtenus sur l'axe avec le détecteur Stingray pour les cibles (a) « Fleurs », (b) « Gauss » (c) « Parterre » (d) « Carreaux ».

Notons qu'avec le détecteur Stingray nous obtenons également un gain par rapport aux résultats expérimentaux obtenus avec l'imageur chromatique conçu à l'IOGS, présenté à la figure 4.2(b), qui ont été obtenus avec ce même détecteur.

En conclusion, avec le détecteur Stingray, nous sommes capables d'estimer la profondeur entre 1 et 5 m avec un biais négligeable et un écart-type proche du pas de calibrage à savoir 5 cm. C'est donc la combinaison du détecteur Stingray et de l'objectif chromatique co-conçu qui constitue pour nous l'imageur CAM3D.

4.4.4 Bilan

Les résultats obtenus avec le détecteur Dalsa présentent un biais important entre 2 et 4 m, qui n'apparaît pas sur les résultats obtenus avec le détecteur Stingray. Il s'agit notamment de la région où les FEP sont les plus petites, donc la région où le repliement est le plus élevé. Ainsi, le repliement spectral semble avoir des conséquences importantes sur la précision d'estimation de profondeur, conséquences qui ne sont pas prévues par notre modèle de performance, il faut donc prendre la précaution de bien dimensionner l'imageur pour limiter ce phénomène.

En conclusion, l'optimisation des plans de mise au point RVB en utilisant le modèle de performance que nous avons développé a permis d'obtenir des résultats d'estimation de profondeur très satisfaisants dans la région définie par le cahier des charges, à savoir entre 1 et 5 m, avec le détecteur Stingray, qui constituent les meilleurs résultats d'estimation de

Figure 4.19	(a)	(b)	(c)	(d)
Biais moyen (valeur absolue)(cm)	4	4	10	4
Écart-type moyen (cm)	7	6	7	6

Tableau 4.6: Performances d'estimation de profondeur moyennes sur la plage de profondeurs comprises entre 1 et 5 m pour les quatre cibles de la figure 4.19. Le biais moyen est calculé en moyennant la valeur absolue des biais expérimentaux.

profondeur obtenus dans cette thèse. Ces résultats illustrent ainsi l'intérêt de la démarche de co-conception.

4.4.5 Cartes de profondeur

La figure 4.20 présente deux images acquises avec l'imageur CAM3D et les cartes de profondeur obtenues avec l'algorithme SAC-GLDFD et les FEP calibrées hors axe dans 9 régions de l'espace, comme nous l'avons fait dans le chapitre 2. Nous présentons également les cartes de profondeur obtenues avec la Kinect.

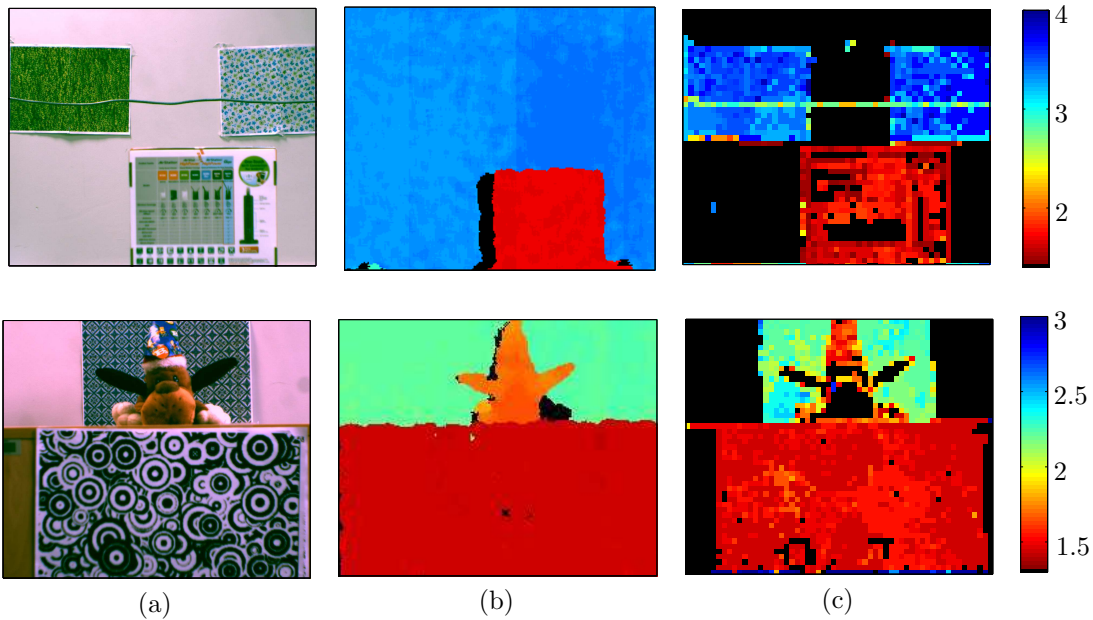


Fig. 4.20: Exemples de cartes de profondeur obtenues avec l'imageur CAM3D. L'échelle des profondeurs est en m. (a) Image RVB. (b) Carte de profondeur produite par la Kinect. (c) Carte de profondeur obtenues avec l'algorithme SAC-GLDFD.

Nous pouvons observer que les cartes de profondeur produites par l'imageur CAM3D et l'algorithme SAC-GLDFD sont cohérentes avec les cartes de profondeurs produites par la Kinect. Par contre, sur la première image, le câble n'a pas été détecté par la Kinect alors qu'il apparaît nettement sur la carte de profondeur produite en utilisant l'imageur CAM3D. Des discontinuités sur la profondeur apparaissent cependant à cause de la séparation de l'image en zones de FEP constante.

4.5 Discussion

Dans ce chapitre, nous avons présenté un procédé de conception d'un imageur chromatique à capacité 3D illustré par un premier exemple de réalisation qui a été validé par la réalisation concrète d'un imageur chromatique en fonction d'un cahier des charges réaliste. Nous discutons ici des améliorations que nous pouvons apporter compte tenu de cette première expérience.

4.5.1 Introduction des critères de performance dans la fonction de mérite de Zemax

Lors de la conception de l'imageur CAM3D, les critères de performance n'ont pas été directement introduits dans la fonction de mérite du logiciel de conception optique et nous avons manuellement optimisé plusieurs imageurs en fonction du chromatisme et de la position du plan de mise au point, afin de sélectionner une configuration favorable vis-à-vis de nos critères de performance. Or il est a priori possible d'introduire des critères définis par l'utilisateur dans la fonction de mérite de Zemax et la bibliothèque zdde permet d'optimiser l'imageur à partir de Matlab. Ceci demande cependant une étape de programmation qui n'a pas été réalisée au cours de cette thèse, à cause du délai que nous avons fixé pour concevoir l'imageur. L'introduction de ces critères directement dans la fonction de mérite de Zemax permettrait d'optimiser automatiquement l'architecture optique en prenant en compte les critères de performance caractérisant la précision d'estimation de profondeur et la capacité à produire une image de grande profondeur de champ. Ce type de fonctionnalité permet d'envisager le développement d'un procédé d'aide à la conception d'un imageur à capacité 3D, qui est décrit dans le brevet que nous avons déposé.

4.5.2 Optimisation de la forme de la pupille codée

Dans la conception de l'imageur CAM3D, nous avons mis de côté l'optimisation d'une pupille codée, afin de simplifier l'optimisation comportant déjà une combinatoire importante. Nous avons cependant laissé accessible la pupille afin de pouvoir faire une optimisation a posteriori de sa forme pour améliorer les précisions d'estimation de profondeur. Dans une première étude rapide nous avons comparé la performance d'estimation de profondeur de l'imageur CAM3D pour plusieurs formes de pupilles. La figure 4.21 présente les valeurs de σ_{BCR} obtenues pour trois pupilles : la pupille circulaire classique et deux autres pupilles présentées à la figure 4.22, pour des profondeurs comprises entre 1 et 10 m. Les FEP simulées sous Zemax à 8 m pour les trois pupilles sont représentées à la figure 4.23.

La figure 4.21 illustre que le gain en précision apporté par les pupilles non circulaires apparaît surtout pour des profondeurs proches ou éloignées des plans de mise au point RVB. Ce phénomène peut s'expliquer par le fait que la forme de la pupille se distingue dans la FEP uniquement pour de fortes défocalisations, comme l'illustre la figure 4.23, ce qui correspond aux profondeurs proches et éloignées des plans de mise au point RVB. Dans cet exemple, la pupille annulaire est celle qui améliore le plus la précision d'estimation de profondeur.

Notre intuition est donc que la combinaison d'une pupille codée et d'un imageur chromatique, n'améliore pas la performance d'estimation de profondeur pour des profondeurs proches des plans de mise au point RVB mais peut augmenter la plage de profondeur où l'estimation est possible en améliorant la précision lorsque la défocalisation est importante. Il serait intéressant de réaliser une pupille annulaire et de caractériser la précision de l'imageur CAM3D avec cette pupille afin de voir si expérimentalement cette dernière améliore les

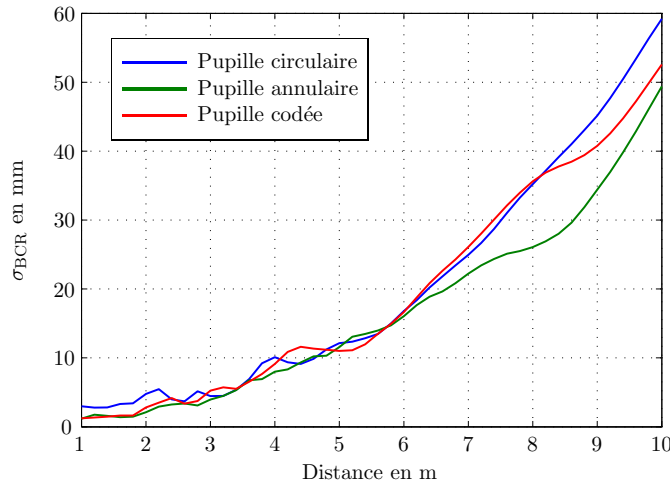


Fig. 4.21: Performances théoriques d'estimation de profondeur de l'imageur CAM3D pour différentes formes de pupilles.

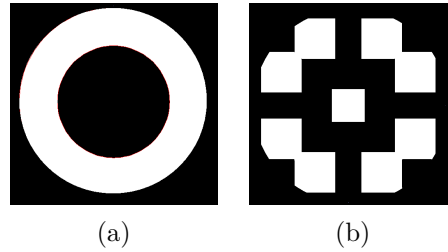


Fig. 4.22: (a) Pupille annulaire. (b) Pupille codée.

performances d'estimation de profondeur après 5 m et entre 1 et 2 m. De plus, une étude plus exhaustive des formes de pupilles potentielles devrait être menée pour déterminer des formes éventuellement plus performantes.

Notons par ailleurs qu'à terme, nous aurions intérêt à optimiser plus tôt la forme de la pupille, notamment au cours de la conception de l'imageur, en construisant une famille de pupilles codées potentielles et en les introduisant dans le logiciel de conception optique lors de l'optimisation.

4.5.3 Variation de la FEP dans le champ

Une des difficultés liées à l'estimation de la profondeur dans l'image entière concerne la variation de la FEP dans le champ. En effet, cette variation implique soit des erreurs d'estimation de profondeur lorsque la FEP utilisée est la FEP calibrée sur l'axe, soit une démarche complexe de calibrage dans laquelle l'image est séparée en différentes régions où la FEP est supposée constante. Il serait donc intéressant d'ajouter une contrainte au moment de l'optimisation de l'imageur afin de favoriser une FEP invariante dans le champ. Cependant, le risque est d'avoir à ajouter de nouvelles lentilles ce qui augmente la complexité et le coût de l'imageur. Par ailleurs, une méthode d'interpolation de FEP à partir d'un nombre restreint de FEP calibrées hors axe permettrait d'éviter le calibrage de la FEP dans le champ pour toutes les profondeurs, ce qui est relativement fastidieux.

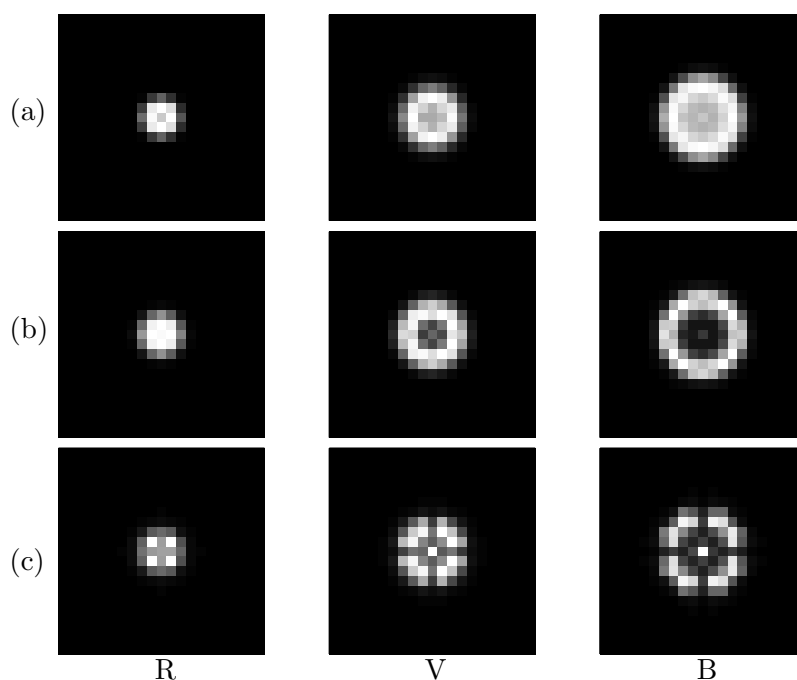


Fig. 4.23: FEP des canaux RVB à 8 m pour différentes formes de pupilles. (a) Pupille circulaire. (b) Pupille annulaire. (c) Pupille codée.

4.5.4 Limitation du repliement

La comparaison des performances d'estimation de profondeur avec les deux détecteurs a montré l'apparition d'un biais pour le détecteur présentant le pas d'échantillonnage le plus grand. Nous expliquons ce biais par des effets de repliement spectraux qui perturbent les données lors du traitement. Cependant, le modèle de performance que nous utilisons ne permet pas de prévoir ce biais. Il est donc nécessaire de prendre en compte les effets de repliement lors du pré-dimensionnement de l'imageur.

Une manière de limiter l'effet du repliement est de faire en sorte que la FEP de diffraction de l'optique soit de dimension voisine de la taille du détecteur. Cela n'était notamment pas le cas pour le détecteur Dalsa. En effet, la tache d'Airy de l'objectif co-conçu a une largeur de $4 \mu\text{m}$ ce qui est bien inférieur à la taille du pixel du détecteur Dalsa qui est de $7.4 \mu\text{m}$, ce qui implique du repliement lié à l'échantillonnage au niveau des plans de mise au point de chaque canal. En contrepartie, comme nous avons pu le voir à la figure 4.11, la qualité image du détecteur Stingray est moins bonne au niveau des plans de mise au point, ce qui limite la capacité à produire une image RVB de grande profondeur de champ. Il y a donc à nouveau un compromis à trouver entre la qualité image et la précision d'estimation de profondeur.

4.5.5 Un pas de calibrage plus fin

Dans les résultats que nous avons obtenu à la figure 4.19, l'écart-type expérimental est très proche du pas de calibrage. Ainsi, il est possible que la précision expérimentale actuelle soit limitée par le pas de calibrage et qu'un pas plus fin permettrait de mieux mesurer la précision d'estimation de profondeur. Le DOTA dispose d'un banc optique long d'environ deux mètres sur lequel est placée une platine qu'il est possible de déplacer automatiquement avec une précision inférieure au mm. Le montage de l'imageur CAM3D sur cette platine

permettrait de calibrer automatiquement et avec un pas très fin les FEP et d'affiner la caractérisation de la performance d'estimation de profondeur. Ce procédé permettrait à la fois un gain de temps de calibrage et une estimation plus fine de la précision d'estimation de profondeur.

4.6 Conclusion

Dans ce chapitre, nous avons présenté une démarche de co-conception d'un imageur chromatique à capacité 3D. D'une manière générale, la conception conjointe repose sur l'ajout de nouveaux critères de performances aux critères caractérisant la qualité image utilisés traditionnellement. Dans notre cas ces nouveaux critères sont la précision d'estimation de profondeur et la capacité à produire une image RVB de grande profondeur de champ. Nous avons repris la démarche classique consistant à prédimensionner l'imageur à partir d'un cahier des charges, puis à concevoir une première architecture à partir des caractéristiques de l'imageur obtenues avec ce prédimensionnement et enfin à optimiser cette architecture au sein du logiciel de conception optique. La différence est que notre conception repose sur de nouveaux critères de performance relatifs à l'estimation de profondeur et à la production d'une image RVB de grande profondeur de champ. À l'issue d'une réflexion originale, nous avons proposé de fonder le prédimensionnement sur les paramètres optiques classiques tels que l'ouverture, la distance focale et les dimensions du détecteur, auxquels nous avons ajouté les positions des plans de mise au point RVB. Nous avons montré que cet ensemble de paramètres permet de caractériser au premier ordre la performance de l'imageur à capacité 3D chromatique et permet à un opticien de faire le choix de l'architecture optique.

Grâce à une collaboration avec le DOTA, nous avons pu mettre en œuvre concrètement cette démarche lors de la conception conjointe de l'imageur CAM3D. Partant d'un cahier des charges associé à la navigation des micro-drones, le prédimensionnement a fourni les valeurs approximatives des paramètres du système qui ont servi de point de départ pour la conception de l'architecture de l'imageur. Nous avons ensuite modifié les plans de mise au point de cette architecture pour optimiser les critères de performance choisis en jouant sur le chromatisme longitudinal et sur la position du détecteur de l'objectif chromatique.

Un des résultats de cette étude est que la position optimale des plans de mise au point RVB pour maximiser la précision diffère de la position optimale de ces plans pour la restauration d'une image RVB par un traitement de type transfert de fréquences. Nous avons donc eu à faire un compromis entre ces deux critères. La réalisation de l'architecture ainsi optimisée a permis d'obtenir le premier prototype d'imageur chromatique co-conçu pour des applications d'estimation de profondeur.

La précision d'estimation de profondeur expérimentale de l'objectif chromatique co-conçu a été évaluée sur deux détecteurs possédant des pixels de taille différente. Nous avons ainsi pu souligner l'influence du repliement spectral sur les performances d'estimation de profondeur. Pour le détecteur possédant les plus grands pixels, ce phénomène se traduit par l'apparition d'un biais sur les résultats. Par contre, l'objectif chromatique co-conçu combiné avec le détecteur possédant les pixels les plus petits, a permis d'obtenir des performances d'estimation de profondeur sur l'axe très satisfaisantes, avec un biais négligeable et un écart-type de l'ordre du pas de calibrage, à savoir 5 cm, sur toute la plage de profondeur comprise entre 1 et 5 m, et ce pour une grande variété de cibles. Avec ce prototype, que nous désignons par CAM3D, nous atteignons la meilleure précision sur la gamme de profondeur la plus large parmi toutes les configurations étudiées dans le contexte de cette thèse, ce

résultat illustre donc l'intérêt de la démarche de co-conception.

Concernant les perspectives de ce travail, notre objectif est d'améliorer le procédé d'optimisation mis en œuvre dans ce chapitre en introduisant les critères de performances directement au sein du logiciel de conception optique afin d'optimiser les paramètres de l'objectif chromatique de manière automatique. Si l'ajout d'une pupille codée ne semble pas influencer l'estimation de profondeur pour des profondeurs proches des plans de mise au point RVB, elle conduit à une amélioration de la précision lorsque les trois canaux ont un flou important. Ce résultat indique que l'utilisation d'une pupille codée pourrait permettre d'agrandir la région de l'espace où la profondeur peut être estimée précisément. Pour étudier cette possibilité, il conviendrait d'introduire la pupille codée dès le début de la conception, en intégrant un paramétrage de sa forme dans les paramètres sur lesquels porte l'optimisation.

À moyen terme, nous souhaitons concevoir un imageur dont les dimensions soit suffisamment réduites afin d'envisager l'intégration sur un micro-drone. Pour cela, un nouveau travail de conception est nécessaire, prenant en considération en plus des performances d'estimation de profondeur, des contraintes d'encombrement et de poids de l'imageur.

Conclusion générale et perspectives

Conclusion générale

Dans cette thèse nous avons élaboré une approche de co-conception, appliquée à la réalisation d'un imageur monovoie passif à capacité 3D. Ce type d'imageur estime la profondeur de la scène observée en utilisant le principe de *Depth from Defocus* (DFD) qui repose sur une estimation locale du flou de défocalisation.

Nous avons tout d'abord développé un nouvel algorithme d'estimation de profondeur suffisamment général pour s'appliquer à une large famille d'imageurs monovoies passifs. Cet algorithme appelé GLDFD, pour *Generalized Likelihood Depth From Defocus*, sélectionne une FEP parmi une famille de FEP potentielles à l'aide d'un critère de maximum de vraisemblance noté GL, pour *Generalized Likelihood*, obtenu par marginalisation de la scène. L'algorithme GLDFD est non supervisé et nous avons proposé une implémentation rapide et efficace adaptée au traitement de fenêtres de taille réduite de l'image. Cet algorithme a montré de bonnes performances en termes de robustesse à la scène et au bruit, sur une variété de familles de FEP potentielles et sur des images simulées et réelles. Enfin, le formalisme utilisé pour développer cet algorithme est facilement généralisable à l'estimation de profondeur multi-images.

Grâce à une collaboration avec des enseignants de l'IOGS, nous avons ensuite étudié deux concepts optiques favorisant l'estimation de profondeur : l'introduction d'une pupille codée et l'utilisation d'une optique, dite chromatique, dont l'aberration chromatique longitudinale n'est pas corrigée. Nous avons dans un premier temps évalué expérimentalement la performance d'estimation de profondeur d'un imageur à pupille codée, en utilisant l'algorithme GLDFD, et montré qu'il était possible de générer des cartes de profondeur cohérentes avec celles produites par un dispositif actif d'estimation de profondeur. Puis nous avons réalisé un imageur chromatique à capacité 3D, ce qui a demandé de développer un nouvel algorithme d'estimation de profondeur, nommé SAC-GLDFD, pour *Single Acquisition Chromatic Depth from Defocus*, qui repose sur la minimisation d'un critère GL_{SAC} proche du critère GL mais construit à partir d'une modélisation de la corrélation entre les canaux RVB de l'imageur chromatique. Un prototype d'imageur chromatique a été réalisé dont nous avons caractérisé expérimentalement la précision. Cette étude a permis d'illustrer l'intérêt de la solution chromatique en termes de précision et de région de l'espace où la profondeur est estimée vis-à-vis de l'estimation de profondeur avec un imageur conventionnel ou à pupille codée.

Nous avons ensuite développé un modèle de performance original, permettant de prévoir la précision d'estimation de profondeur d'un imageur en fonction des paramètres de l'optique et des traitements dans les cas de l'acquisition d'une seule image (SA-DFD, pour *Single Acquisition DFD*), de plusieurs images de la même scène (MA-DFD, pour *Multiple Acquisitions DFD*) et l'acquisition d'images par un imageur chromatique (SAC-DFD, pour *Single Acquisition Chromatic DFD*). Une étude approfondie du cas SA-DFD a permis de mettre en évidence les paramètres influençant cette précision et nous avons montré que pour des distances éloignées du plan de mise au point, diminuer la taille de la FEP permet d'améliorer la précision d'estimation de profondeur. Pour chacune des configurations nous avons illustré le caractère prédictif de notre modèle en vérifiant que la configuration de l'imageur conduisant à la meilleure précision théorique était identique à la configuration la plus précise expérimentalement. Les outils développés permettent alors de comparer les performances théoriques d'une large gamme d'imageurs, ce qui permet d'aborder la co-conception proprement dite.

Ces outils de modélisation de performance ont été appliqués à la conception d'un imageur chromatique à capacité 3D, appelé CAM3D, au cours d'un travail mené en collaboration avec le Département d'Optique Théorique et Appliquée de l'Onera. Nous avons défini un cahier des charges, qui impose la région de l'espace dans laquelle l'imageur doit pouvoir estimer la profondeur et fournir après traitement une image nette ainsi que la résolution de la carte de profondeur et les conditions d'illumination. Nous avons ensuite effectué un prédimensionnement des principaux paramètres de l'imageur. Ce prédimensionnement avait pour objectif de fixer la distance focale, l'ouverture, les dimensions du détecteur et la position des plans de mise au point RVB, paramètres dont nous avons montré qu'ils permettent d'avoir un ordre de grandeur de la performance de l'imageur. Ce dimensionnement a été réalisé à partir d'une modélisation très simple de l'imageur et des FEP et en utilisant un critère de performance d'estimation de profondeur combiné à un critère caractérisant la capacité à produire une image de grande profondeur de champ. Puis, à partir d'une architecture vérifiant les caractéristiques issues du prédimensionnement, nous avons affiné les caractéristiques de l'imageur à l'aide de nos deux critères de performance. Nous avons montré qu'il faut alors choisir un compromis entre précision d'estimation de profondeur et qualité de l'image restaurée. Nous avons ensuite fait réaliser un prototype de l'objectif chromatique à partir des spécifications obtenues. Ce prototype a été évalué expérimentalement sur deux détecteurs possédant des tailles de pixels différents. Le premier détecteur, sujet au repliement spectral à cause de la taille importante de ses pixels, présente un biais non négligeable dans les régions de profondeurs proches des plans de mise au point RVB, ce qui a illustré l'influence de ce phénomène sur la performance d'estimation de profondeur. Les résultats d'estimation de profondeur obtenus à l'aide du deuxième détecteur possédant de plus petits pixels, et donc moins sujet au repliement, ont permis d'atteindre la meilleure précision sur la gamme de profondeur la plus large parmi toutes les configurations étudiées dans le contexte de cette thèse, ce qui illustre l'intérêt de la co-conception.

Perspectives

Amélioration des traitements

Accélération des algorithmes

Les algorithmes d'estimation de profondeur développés sous **Matlab** au cours de cette thèse n'ont pas spécifiquement été optimisés en termes de temps de calcul. Un travail d'accélération de ces traitements est en cours au Département de Traitement de l'Information et

Modélisation de l'Onera (DTIM), notamment à l'aide d'une parallélisation en GPU. En effet, puisque les critères GL et GL_{SAC} sont calculés de la même façon pour chaque fenêtre de l'image et pour chaque FEP, les algorithmes GLDFD et SAC-GLDFD se prêtent bien à la parallélisation. Lors de son stage de fin d'études, S. Vittorino s'est intéressée à l'implémentation de l'algorithme GLDFD sous OpenCL. Les temps de calcul obtenus pour les images issues de [Levin *et al.*, 2007b] et 9 FEP sont inférieurs à la seconde sachant que le temps de calcul initial était de 1 min. La suite de ce travail consiste à présent à développer une version parallélisée de l'algorithme SAC-GLDFD, travail en cours en collaboration avec A. Plyer. Outre la parallélisation des algorithmes, la démarche de sélection des FEP peut également être optimisée pour diminuer les temps de calcul. En effet, en estimant la profondeur avec un pas très important puis en affinant la recherche avec des FEP dans un intervalle de profondeur plus restreint, il est possible de réduire le nombre de calculs à effectuer pour chaque fenêtre. Si pour le moment la cadence vidéo n'est pas encore atteinte, les temps de calcul des cartes de profondeur ont été significativement diminués.

Amélioration du modèle a priori

L'a priori gaussien utilisé dans les différents algorithmes GLDFD et SAC-GLDFD suppose respectivement que les gradients de la scène en niveau de gris, ou ceux de la luminance et des chrominances, ont une distribution isotrope. Or pour des situations réelles, des régions d'une image peuvent présenter des valeurs de gradients plus importantes dans une direction particulière. Dans un premier temps, nous pourrions considérer des distributions gaussiennes localement anisotropes dont les caractéristiques directionnelles seraient estimées directement sur l'image. Une approche plus ambitieuse consisterait à adapter les techniques développées pour la déconvolution aveugle à l'aide de modèles non gaussiens [Levin *et al.*, 2007b].

Régularisation des cartes de profondeur

Les cartes de profondeurs que nous avons présentées dans cette thèse correspondent aux résultats bruts des algorithmes GLDFD et SAC-GLDFD. Or, nous avons constaté la présence sur ces résultats de valeurs aberrantes, et de plus la profondeur n'est pas estimée sur les régions homogènes. Un traitement de régularisation peut permettre d'éliminer les valeurs aberrantes de ces résultats et de propager l'information de profondeur dans les régions homogènes. Il existe de nombreuses approches de régularisation, nous en présentons brièvement trois, à savoir l'utilisation d'un filtre médian, d'une technique de segmentation markovienne, ou la fusion de la carte de profondeur avec une segmentation de l'image couleur.

Un premier traitement de régularisation permettant de supprimer des valeurs aberrantes de profondeur consiste à appliquer un filtrage médian sur la carte de profondeur. Le filtre médian ne permet cependant pas de propager l'information de profondeur dans les régions homogènes de l'image. La figure 4.24(d) présente un exemple de résultat obtenu avec un filtrage médian. Ce filtrage permet bien d'éliminer une grande partie des valeurs aberrantes et diminue l'aspect pixellisé de la carte de profondeur, lié au traitement de fenêtres de l'image. En contrepartie il a tendance à épaissir les contours des objets de la scène.

Un autre type de traitement de régularisation consiste à utiliser une méthode de segmentation markovienne suivant le modèle de Potts. Cette méthode est très utilisée dans le domaine de la vision par ordinateur, sous la dénomination *Graphcut* [Wang *et al.*, 2001; Kolmogorov et Zabih, 2004]. Une version de ce traitement a été développée au DTIM par M. Sanfourche et G. Le Besnerais. Cette technique permet d'optimiser un terme de données séparable non convexe voir multimodal. Ce type de régularisation a été notamment proposé

dans [Levin *et al.*, 2007b; Martinello *et al.*, 2010; Martinello et Favaro, 2011b; Chakrabarti *et al.*, 2010; Chakrabarti et Zickler, 2012]. La figure 4.24(e) présente un résultat de cette technique sur les valeurs du critère GL associées à l'image de la figure 4.24(a). Le résultat est beaucoup plus homogène qu'avec un filtre médian et la silhouette est globalement correctement segmentée. Cependant, le fond a été lissé à une seule profondeur.

Enfin, une troisième méthode, développée au DTIM par V. Leung, consiste à fusionner la carte de profondeur avec l'image segmentée en utilisant les informations de couleur de l'image. La segmentation de l'image est réalisée en utilisant l'algorithme de Felzenszwalb et Hottelocher dans l'espace HSV [Felzenszwalb et Huttenlocher, 2004]. Le principe de la fusion de l'image segmentée et de la carte de profondeur est de sélectionner dans chaque région segmentée le label ayant la plus grande occurrence. La figure 4.24(e) présente un résultat de fusion. Nous constatons que cette méthode conduit à une bien meilleure segmentation de la silhouette, par exemple la tête et les bras sont correctement restitués. En revanche la méthode ne parvient pas à propager l'information dans certaines régions du fond et la silhouette a été associée à la même profondeur que le journal.

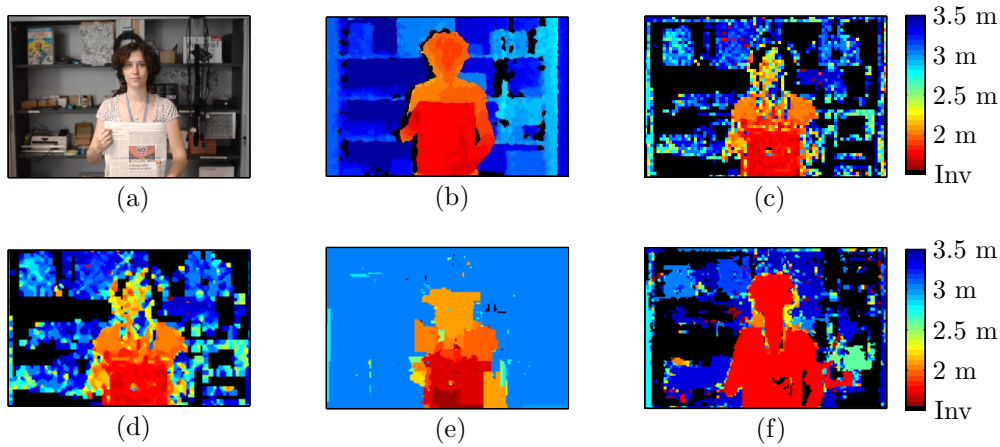


Fig. 4.24: Résultats de différentes approches de régularisation sur des images produites par un imageur à pupille codée. (a) Image RVB. (b) Carte de profondeur produite par la Kinect. (c) Carte de profondeur brute obtenue avec l'algorithme GLDFD. (d) Filtrage médian. (e) *Graphcut* (f) Fusion.

Nous avons considéré ici trois approches possibles de régularisation. En pratique, le type de traitement régularisation va dépendre de l'application et de la priorité donnée au temps de calcul. Parmi les trois méthodes que nous avons testées, seule le filtrage médian est compatible avec les applications de type capteur embarqués pour lesquelles la ressource de calcul est très limitée. De plus, pour ce type d'application, il est plus important de détecter les zones où la profondeur ne peut pas être estimée de façon fiable que d'interpoler ces zones. Dans un contexte où la puissance de calcul disponible n'est pas trop limitée, il est possible d'obtenir des résultats plus précis avec des méthodes plus coûteuses. Les premiers tests que nous avons menés indiquent qu'une approche exploitant une fusion avec les informations de couleur semble la plus prometteuse, ce qui confirme les résultats de divers travaux qui utilisent eux aussi ce type de fusion. Les méthodes précédentes ont été testées sur des images issues d'un imageur à pupille codée, notre objectif est à présent de les tester sur des images issues d'un imageur chromatique.

Restauration d'une image RVB

L'utilisation du chromatisme dégrade la qualité de l'image acquise par l'imageur du fait de la résolution variable en fonction des canaux R,V ou B. Cependant, nous avons vu à la fin du chapitre 2 que la qualité de l'image peut être restaurée par traitement. Comme pour la régularisation des cartes de profondeur, les algorithmes de restauration de l'image RVB sont différents selon l'application recherchée. Les traitements de transfert de fréquence décrits au chapitre 2 sont particulièrement rapides et adaptés pour des applications de drones embarqués. Il reste néanmoins un travail à faire pour optimiser les paramètres de ce transfert. Pour des applications où le temps de calcul est moins critique, nous pouvons envisager des algorithmes de restauration multi-canal globale à partir de FEP localement variables en s'inspirant notamment des méthodes déconvolution multi-canal, et des méthodes de dématricage [Galatsanos *et al.*, 2000; Gotoh et Okutomi, 2004; Condat, 2009].

Co-conception d'un imageur à capacité 3D

Amélioration du procédé de conception

Lors de la conception de l'imageur CAM3D, nous avons utilisé le logiciel de conception optique **Zemax** afin de générer les FEP d'un imageur et évaluer sa précision théorique d'estimation de profondeur à l'aide du modèle que nous avons développé. Cependant, la comparaison des solutions potentielles a été réalisée de manière externe à ce logiciel et de manière supervisée. Or il serait possible, par un développement plus important du dialogue entre **Matlab** et de **Zemax**, de créer une variable supplémentaire dans la fonction de mérite de **Zemax** faisant appel, à chaque étape de l'optimisation de l'imageur par le logiciel, au calcul de la précision d'estimation de profondeur dans une plage de profondeur donnée. Ceci permettrait d'intégrer directement le critère de précision dans les algorithmes d'optimisation du logiciel de conception optique. De même, il serait possible d'introduire une estimation de l'ensemble des profondeurs pour lesquelles un transfert de fréquence est possible, en considérant par exemple les valeurs de FTM en intensité des trois canaux en fonction de la profondeur. L'outil obtenu permettrait d'optimiser un imageur, à partir d'une architecture initiale, compte tenu de contraintes fixées par l'utilisateur telles que le champ, la plage de distance dans laquelle la profondeur doit être estimée et les conditions d'illumination de la scène. L'idée de ce procédé de conception fait l'objet du brevet déposé au cours de ces travaux de thèse.

Étude théorique approfondie des concepts optiques favorisant l'estimation de profondeur

Lors de la conception de CAM3D, nous avons mis de côté l'optimisation d'une pupille codée pour nous consacrer à l'optimisation des paramètres d'un imageur chromatique à capacité 3D et en particulier la position des plans de mise au point RVB. Cependant, il serait intéressant d'utiliser les modèles de performances que nous avons développés pour étudier plus précisément l'influence d'une pupille codée sur la précision d'estimation de profondeur. Cette étude pourrait être menée dans un premier temps dans le cas mono-image SA-DFD, afin de trouver une forme de pupille optimale suivant notre modèle de performance. Nous pourrions alors la comparer expérimentalement aux différentes pupilles proposées dans la littérature. Par ailleurs, il serait intéressant d'approfondir l'étude préliminaire réalisée à la fin du chapitre 4, concernant l'apport d'une pupille codée sur un imageur chromatique. Dans ces études théoriques, il faudrait notamment prendre en compte la diminution de rapport signal à bruit liée à l'utilisation d'une pupille codée en amplitude.

À terme, il serait intéressant d'utiliser nos différents modèles de performances pour conduire des comparaisons de performance théorique entre les solutions optiques du type imageur achromatique, imageur achromatique muni d'une pupille codée, imageur chromatique et imageur chromatique muni d'une pupille codée et de définir un contexte permettant une comparaison juste de ces différentes solutions optiques.

Enfin, d'autres solutions optiques peuvent également être envisagées pour favoriser l'estimation de profondeur, comme par exemple l'utilisation d'un masque de phase.

Nouvelles applications pour la méthode de conception conjointe

L'évaluation expérimentale de la performance d'estimation de profondeur de CAM3D a permis d'illustrer l'intérêt de la conception conjointe, cependant ses dimensions et son poids ne permettent pas d'envisager son intégration sur un micro-drone. À partir de notre première expérience de co-conception, une nouvelle conception d'imageur peut être envisagée, en introduisant de nouvelles contraintes relatives à la dimension et au poids de l'imageur.

De plus, à l'aide du procédé de conception que nous avons développé, il est possible de concevoir de nouvelles générations d'imageurs pour des applications différentes que celle envisagée dans cette thèse, comme par exemple des applications en vélocimétrie par imagerie de particule (PIV). En effet, dans ces domaines la technique d'estimation de profondeur monovoie passive est attractive car elle limite l'encombrement et elle n'est pas intrusive.

Il existe également de multiples domaines autre que la photographie, dans lesquels il est intéressant d'utiliser une approche de conception conjointe. Par exemple en microscopie la forme de la FEP est parfois modifiée pour favoriser l'estimation de la position des particules, par exemple en introduisant de l'astigmatisme [Cierpka *et al.*, 2010], ou en utilisant une FEP à double hélice qui présente deux modes effectuant une rotation en fonction de la profondeur [Greengard *et al.*, 2006; Pavani *et al.*, 2009]. Dans ces deux cas, il serait intéressant d'étudier l'interaction entre les algorithmes de localisation des particules et la forme de la FEP. Par ailleurs, la démarche de co-conception pourrait être employée dans le cas des caméras plénoptiques, pour revenir sur les choix de conception, et les mettre en relation avec des objectifs de qualité image ou d'estimation de profondeur.

Annexes

Dérivation de la vraisemblance dans le cas d'un a priori impropre

Nous présentons ici plusieurs propriétés des vecteurs aléatoires gaussiens impropres. Ces propriétés et leurs démonstration proviennent d'un rapport technique interne de l'ONERA de F. Champagnat [Champagnat, 2012]. Les démonstrations de ces propriétés sont simplement rappelées ici car elles sont appliquées dans ce manuscrit dans le cas particulier de l'estimation de profondeur. Ce travail s'inspire des références [Wahba, 1985; Neumaier, 1998]. En effet, Wahba utilise un a priori gaussien impropre pour l'optimisation du paramètre de régularisation dans un contexte de lissage par spline. Elle développe alors une expression de la vraisemblance calculée en enlevant la valeur propre nulle du déterminant apparaissant dans l'expression de la vraisemblance. Plus récemment, ce procédé est étendu dans [Neumaier, 1998] dans le contexte des problèmes linéaires mal posés. L'approche commune à Wahba et Neumaier consiste à projeter les données impropres sur un sous-espace de dimension inférieure, dans lequel les données une fois projetées ont une matrice de précision régulière, ce qui permet d'utiliser les outils statistiques classiques. Cependant, leurs démonstrations s'appliquent uniquement au cas de l'estimation du paramètre de régularisation, et dans leurs approches, le sous-espace choisi dépend alors de la FEP. Or dans le cas de l'estimation de la profondeur, la FEP fait partie des paramètres à estimer et il faut que la projection des données soit indépendante de la FEP. Les résultats présentés dans cette annexe décrivent l'hypothèse suivant laquelle l'approche de Neumaier et de Wahba peut être généralisée pour définir la vraisemblance d'un vecteur aléatoire gaussien impropre. Ces résultats sont utilisés dans le chapitre 1 pour l'application de l'estimation de la FEP avec un a priori impropre sur la scène.

A.1 Définition d'un VIG

On désigne par VIG les vecteurs aléatoires gaussiens ayant une matrice de précision singulière. Ces vecteurs ne définissent pas une mesure de probabilité mais plutôt une mesure non bornée [Halmos, 1969]. Ceci conduit à la définition plus concrète d'un VIG :

Définition 1 *Un sous espace aléatoire impropre gaussien (VIG), de matrice de précision Q est un vecteur dont la densité s'écrit*

$$p_{\mathbf{X}}(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\mathbf{x}^t Q \mathbf{x}\right) \quad (\text{A.1})$$

où Q est une matrice singulière définie non négative.

Le fait que Q soit singulière empêche la dérivation de la formule classique d'une vraisemblance pour \mathbf{x} telle que celle présentée en (1.34). Nous présentons dans la suite un ensemble de propriétés permettant de contourner cette difficulté.

A.2 Propriétés des VIG par transformation linéaire

Lemme 1 *Soit \mathbf{X} un VIG de matrice de précision Q et A une matrice régulière. Alors $\mathbf{Y} = A\mathbf{X}$ est un VIG ayant comme matrice de précision $A^{-t}QA^{-1}$.*

Démonstration

Soient B un vecteur quelconque mesurable de \mathbb{R}^N et $\mu_{\mathbf{Y}}$ (resp. $\mu_{\mathbf{X}}$) la mesure de \mathbf{Y} (resp. \mathbf{X}). Alors par définition $\mu_{\mathbf{Y}}(B) = \mu_{\mathbf{X}}(A^{-1}(B))$ et

$$\mu_{\mathbf{X}}(A^{-1}(B)) = \int_{A^{-1}(B)} \exp\left(-\frac{1}{2}\mathbf{x}^t Q \mathbf{x}\right) d\mathbf{x}. \quad (\text{A.2})$$

Par changement de variable $\mathbf{y} = A\mathbf{x}$ sous l'intégrale précédente il vient :

$$\int_{A^{-1}(B)} \exp\left(-\frac{1}{2}\mathbf{x}^t Q \mathbf{x}\right) d\mathbf{x} = \int_B \exp\left(-\frac{1}{2}\mathbf{y}^t A^{-t}QA^{-1}\mathbf{y}\right) \frac{d\mathbf{y}}{|A|}. \quad (\text{A.3})$$

Cela implique que \mathbf{Y} est aussi un VIG avec comme matrice de précision $A^{-t}QA^{-1}$. ■

Lemme 2 *Soit \mathbf{Y} un VIG de matrice de précision Q de rang n , soit A_p (resp. A_i) une matrice de taille $n \times N$ (resp. $(N - n) \times N$) dont les vecteurs lignes balayent l'image (resp. le noyau) de Q . Alors $\mathbf{Y}_i = A_i\mathbf{Y}$ et $\mathbf{Y}_p = A_p\mathbf{Y}$ sont des vecteurs aléatoires indépendants tels que \mathbf{Y}_i est une mesure de Lebesgue et \mathbf{Y}_p est un vecteur aléatoire gaussien de matrice de précision régulière $(A_p^+)^tQA_p^+$ (où $^+$ signifie pseudo inverse).*

Démonstration

Soit $A = (A_p^t A_i^t)^t$, puisque Q est non définie négative, $A_p A_i^t = 0$, d'où $A^{-1} = (A_p^+ A_i^+)$. D'après le lemme 1, $(\mathbf{Y}_p^t \mathbf{Y}_i^t)^t = A\mathbf{Y}$ est un VIG dont la densité s'écrit :

$$p_{\mathbf{Y}_p, \mathbf{Y}_i}(\mathbf{y}_p, \mathbf{y}_i) \propto \frac{1}{|A|} \exp\left(-\frac{1}{2}\mathbf{y}_p^t (A_p^+)^tQA_p^+ \mathbf{y}_p\right).$$

Une telle densité est séparable suivant les deux variables (puisque'il s'agit d'une fonction constante de \mathbf{y}_i), donc les vecteurs \mathbf{Y}_p et \mathbf{Y}_i sont indépendants. Enfin, $(A_p^+)^tQA_p^+$ est régulière puisque A_p balaye l'image de Q . ■

Le lemme 2 rejoint le Lemme 3.1 écrit dans [Besag et Kooperberg, 1995] mais ici il est montré que n'importe quel VIG peut être décomposé en deux composantes, l'une ayant une densité de probabilité propre et l'autre une densité impropre. Autrement dit l'espace des VIG peut être considéré comme un ensemble de vecteurs gaussiens aléatoires ayant une matrice de covariance régulière, auxquels s'ajoutent des vecteurs dont la covariance tend vers l'infini.

A.3 Définition d'une vraisemblance associée à un VIG

Proposition 1 Soit \mathbf{Y} un VIG paramétré par la matrice de précision Q_θ . Si le noyau de Q_θ , $\text{Ker } Q_\theta$, ne dépend pas de θ , alors la vraisemblance peut être définie de la manière suivante :

$$L(\mathbf{y}|\theta) = \left| \frac{Q_\theta}{2\pi} \right|_+^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{y}^t Q_\theta \mathbf{y} \right), \quad (\text{A.4})$$

où $|Q_\theta|_+$ est le produit des valeurs propres non nulles de Q_θ .

Démonstration

Première partie

Puisque Q_θ est définie non négative, l'image de Q_θ est orthogonale à son noyau. Ainsi, puisque par définition le noyau de Q_θ ne dépend pas de θ , son image est aussi indépendante de θ . Soit A_p (resp. A_i) une matrice de taille $n \times N$ (resp. $(N - n) \times N$) dont les lignes sont des vecteurs appartenant à l'image (resp. au noyau) de Q_θ , A_p (resp. A_i) est alors indépendante de θ . En utilisant le lemme 2 il est possible de définir une transformation linéaire bijective $\mathbf{Y} = (\mathbf{Y}_p^t \mathbf{Y}_i^t)^t$, où \mathbf{Y}_p et \mathbf{Y}_i sont deux vecteurs aléatoires indépendants tels que $\mathbf{Y}_p = A_p \mathbf{Y}$ et $\mathbf{Y}_i = A_i \mathbf{Y}$. Plus précisément, \mathbf{Y}_i est une mesure de Lebesgue, indépendante de θ et \mathbf{Y}_p est un vecteur gaussien de matrice de précision régulière, $(A_p^+)^t Q_\theta A_p^+$. De ce fait, toute l'information concernant θ est contenue dans \mathbf{Y}_p : selon la terminologie statistique, \mathbf{Y}_p joue donc le rôle de *statistique suffisante*. Ainsi, la vraisemblance de θ connaissant \mathbf{y} peut être définie en utilisant la vraisemblance du paramètre θ vis-à-vis du vecteur \mathbf{Y}_p :

$$L(\mathbf{y}|\theta) = p_{\mathbf{Y}_p}(A_p \mathbf{y}) = \left| \frac{(A_p^+)^t Q_\theta A_p^+}{2\pi} \right|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{y}^t A_p^t (A_p^+)^t Q_\theta A_p^+ A_p \mathbf{y} \right). \quad (\text{A.5})$$

Le terme dans l'exponentielle se simplifie en utilisant :

$$A_p^t (A_p^+)^t Q_\theta A_p^+ A_p = Q_\theta, \quad (\text{A.6})$$

car $A_p^+ A_p$ est une projection orthogonale vers l'image de Q_θ . Cependant, même avec la simplification donnée par l'équation (A.6), l'équation (A.5) n'est pas pratique à utiliser à cause du déterminant qui dépend d'un choix particulier de la base de projection sur l'image de Q_θ .

Deuxième partie.

La matrice A_p peut être choisie telle que ses lignes définissent une base orthonormale de l'image de Q_θ , par un tel choix nous allons voir que :

$$|A_p^{+t} Q_\theta A_p^+| = |Q_\theta|_+. \quad (\text{A.7})$$

Tout d'abord à cause de l'orthonormalité : $A_p^+ = A_p^t$, il faut donc montrer : $|A_p Q_\theta A_p^t| = |Q_\theta|_+$. Par décomposition en valeurs propres/vecteurs propres il vient :

$$Q_\theta = (B_\theta^t B_\theta) \begin{pmatrix} \Lambda_\theta & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} B_\theta \\ B_2 \end{pmatrix} \quad (\text{A.8})$$

$B_\theta^t \Lambda_\theta B_\theta$ constitue donc la décomposition en valeurs propres/ vecteurs propres incluant uniquement les valeurs propres non nulles de Q_θ . Ceci implique :

$$A_p Q_\theta A_p^t = (A_p B_\theta^t) \Lambda_\theta (A_p B_\theta^t)^t. \quad (\text{A.9})$$

Par définition $A_p A_p^t = I$ et $B_\theta B_\theta^t = I$. De plus, nous avons $\text{Ker } A_p = \text{Ker } B_\theta = \text{Ker } Q_\theta$. Par ailleurs d'après le Lemme 3 démontré ci dessous, $A_p B_\theta^t$ est une matrice orthonormale d'où $|A_p B_\theta^t|^2 = 1$. Il s'ensuit $|A_p Q_\theta A_p^t| = |\Lambda_\theta| = |Q_\theta|_+$.

Conclusion

L'équation (A.4) est obtenue en regroupant les équations (A.5), (A.6) et (A.7). ■

Lemme 3 Soient A et B deux matrices de taille $n \times N$ telles que $AA^t = I$, $BB^t = I$ et $\text{Ker } A = \text{Ker } B$. Alors $AB^t BA^t = I$.

Démonstration

Les lignes de A définissent un ensemble de vecteurs orthonormés. Cet ensemble peut être complété par $N - n$ vecteurs orthonormaux afin de former une base orthonormale. Soit C la matrice $(N - n) \times N$ contenant ces vecteurs. Alors $\tilde{A} \triangleq [A^t C^t]^t$ est une matrice unitaire de taille $N \times N$. $\tilde{B} \triangleq [B^t C^t]^t$ l'est également car $BB^t = I$ et $CB^t = 0$ (car $\text{Ker } (A) = \text{Ker } (B)$). Le développement de $\tilde{A}\tilde{B}^t$ implique :

$$\tilde{A}\tilde{B}^t = \begin{pmatrix} AB^t & AC^t \\ CB^t & CC^t \end{pmatrix} = \begin{pmatrix} AB^t & 0 \\ 0 & I \end{pmatrix} \quad (\text{A.10})$$

En tant que produit de matrice unitaire $\tilde{A}\tilde{B}^t$ est également unitaire d'où

$$\tilde{A}\tilde{B}^t \tilde{B}\tilde{A}^t = \begin{pmatrix} AB^t BA^t & 0 \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \quad (\text{A.11})$$

Donc $AB^t BA^t = I$ et ainsi AB^t est une matrice orthonormale. ■

A.4 Cas d'un VIG filtré et bruité

Dans le chapitre 1, nous manipulons des VIG filtrés et bruités. C'est notamment le cas du vecteur de scène \mathbf{x} qui, une fois filtré et bruité, donne le vecteur \mathbf{y} . Pour définir la vraisemblance du vecteur \mathbf{y} , nous pouvons utiliser la proposition 1. Cependant, le point important est de vérifier que le noyau de la matrice de précision paramétrée de \mathbf{y} est indépendant du paramètre. Le lemme 4 démontre une propriété utile sur le noyau d'un vecteur aléatoire filtré et bruité issu d'un VIG pour vérifier cette propriété.

Lemme 4 Soit \mathbf{Y} le vecteur aléatoire défini par $\mathbf{Y} = H\mathbf{X} + \mathbf{B}$ où \mathbf{X} est un VIG de matrice de précision non définie négative Q_x et \mathbf{B} un vecteur aléatoire gaussien ayant une matrice de précision définie positive Q_b . La matrice de précision de \mathbf{Y} notée Q_y s'écrit :

$$Q_y = Q_b - Q_b H (H^t Q_b H + Q_x)^{-1} H^t Q_b. \quad (\text{A.12})$$

Si la matrice H est telle que $H^t Q_b H + Q_x$ est régulière, alors $\text{Ker } Q_y = H \text{Ker } Q_x$.

Démonstration

En effectuant la transformation linéaire $[\mathbf{x}, \mathbf{y}] \rightarrow [\tilde{\mathbf{x}}, \mathbf{y}]$ on a pour tout \mathbf{x}, \mathbf{y} ,

$$\mathbf{x}^t Q_x \mathbf{x} + (\mathbf{y} - H\mathbf{x})^t Q_b (\mathbf{y} - H\mathbf{x}) = \tilde{\mathbf{x}}^t Q_{\tilde{\mathbf{x}}} \tilde{\mathbf{x}} + \mathbf{y}^t Q_y \mathbf{y} \quad (\text{A.13})$$

où $\tilde{\mathbf{x}} = \mathbf{x} - Q_x^{-1} H^t Q_b \mathbf{y}$ et $Q_{\tilde{\mathbf{x}}} = H^t Q_b H + Q_x$.

Montrons d'abord que $\text{Ker } Q_y \subset H \text{Ker } Q_x$. Soit $\mathbf{y} \in \text{Ker } Q_y$, et $\mathbf{x} = Q_x^{-1} H^t Q_b \mathbf{y}$. Pour

de tels \mathbf{x} et \mathbf{y} , $\tilde{x} = 0$ donc le terme de droite de (A.13) est égal à zéro. Ainsi les deux termes de la partie gauche de (A.13) sont nuls. Alors $\mathbf{x} \in \text{Ker } Q_x$ et $\mathbf{y} = H\mathbf{x}$ d'où $\mathbf{y} \in H \text{ Ker } Q_x$.

Montrons ensuite que $H \text{ Ker } Q_x \subset \text{Ker } Q_y$. Soit $\mathbf{y} \in H \text{ Ker } Q_x$, il existe au moins un \mathbf{x} tel que $\mathbf{y} = H\mathbf{x}$ et $Q_x \mathbf{x} = 0$. Pour de tels \mathbf{x} et \mathbf{y} , la partie de gauche de (A.13) est nulle, donc les deux termes de la partie droite de (A.13) sont nuls. En particulier $\mathbf{y}^t Q_y \mathbf{y} = 0$ ce qui signifie que $\mathbf{y} \in \text{Ker } Q_y$. ■

Simulation de FEP à partir des formules d'optique de Fourier

L'objectif de cette annexe est de présenter les formules d'optique physique permettant de modéliser la FEP optique d'un imageur en tenant compte de la défocalisation [Goodman, 1996]. Nous supposons ici que l'optique ne présente aucune autre aberration. Vis-à-vis de la modélisation de la FEP décrite dans la section 1.1.1, nous présentons une modélisation de la fonction h_{opt} .

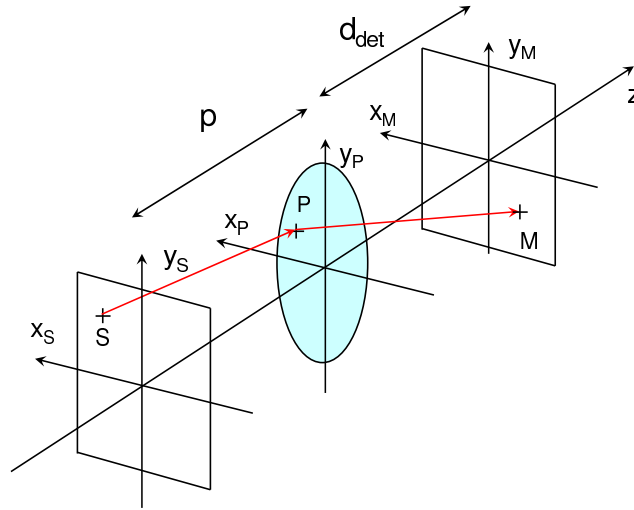


Fig. B.1: Modélisation de la FEP en optique de Fourier.

Soit un point source placé en $S (x_S, y_S, 0)$, schématisé sur la figure B.1, produisant une onde sphérique. Soit P de coordonnées (x_P, y_P, p) un point quelconque de l'espace, l'amplitude complexe au niveau du point P s'écrit :

$$U_{inc}(P) = U_0 \frac{e^{ik\overline{SP}}}{\overline{SP}}, \quad (\text{B.1})$$

avec $k = 2\pi/\lambda$ avec λ la longueur d'onde de l'onde incidente. Nous supposons que cette onde sphérique traverse une lentille de distance focale f , supposée infiniment fine. L'amplitude

complexe juste après la traversée de la lentille s'écrit alors :

$$U(P) = U_{inc}(P)e^{i\phi}, \quad (B.2)$$

où ϕ est la phase introduite par la lentille. Cette onde se propage ensuite entre la lentille et le détecteur. D'après la formule de Rayleigh et Sommerfeld et en se plaçant dans l'approximation paraxiale, l'amplitude d'un point M sur le plan du détecteur s'écrit :

$$U(M) = \frac{-i}{\lambda} \iint A(x_P, y_P) U_0 e^{i\phi} \frac{e^{ik\overline{SP}}}{\overline{SP}} \frac{e^{ik\overline{PM}}}{\overline{PM}} dx_P dy_P. \quad (B.3)$$

Avec $A(x, y)$ la fonction définissant la pupille de la lentille. En utilisant l'approximation de Fresnel, nous pouvons écrire :

$$\overline{SP} \simeq p + \frac{(x_S - x_P)^2 + (y_S - y_P)^2}{2p} \quad (B.4)$$

$$\overline{PM} \simeq d_{det} + \frac{(x_M - x_P)^2 + (y_M - y_P)^2}{2d_{det}}, \quad (B.5)$$

$$(B.6)$$

ainsi que :

$$\phi = \frac{-\pi}{\lambda f} (x_P^2 + y_P^2). \quad (B.7)$$

d'où :

$$U(M) \simeq \frac{-i}{\lambda} \frac{e^{ikp}}{p} \frac{e^{ikd_{det}}}{d_{det}} e^{\frac{ik(x_S^2 + y_S^2)}{2p}} e^{\frac{ik(x_M^2 + y_M^2)}{2d_{det}}} \iint A(x_P, y_P) U_0 e^{-i\frac{\pi}{\lambda} \left(\frac{1}{f} - \frac{1}{p} - \frac{1}{d_{det}} \right) (x_P^2 + y_P^2)} e^{-ik \left(\frac{x_S}{p} + \frac{x_M}{d_{det}} \right) x_P + \left(\frac{y_S}{p} + \frac{y_M}{d_{det}} \right) y_P} dx_P dy_P \quad (B.8)$$

En posant :

$$\phi_{def}(x_P, y_P) = \frac{\pi}{\lambda} \left(\frac{1}{f} - \frac{1}{p} - \frac{1}{d_{det}} \right) (x_P^2 + y_P^2), \quad (B.9)$$

nous faisons apparaître la phase liée au flou de défocalisation qui s'annule lorsque p vérifie la formule de conjugaison de la lentille. La FEP est alors proportionnelle au module carré de l'amplitude complexe en M .

$$h_{opt}(M) \propto |U(M)|^2 \quad (B.10)$$

$$\propto \left| \iint A(x_P, y_P) e^{-i\phi_{def}(x_P, y_P)} e^{-ik \left[\left(\frac{x_S}{p} + \frac{x_M}{d_{det}} \right) x_P + \left(\frac{y_S}{p} + \frac{y_M}{d_{det}} \right) y_P \right]} dx_P dy_P \right|^2 \quad (B.11)$$

Cette expression fait apparaître une transformée de Fourier, centrée au point $x_M = \gamma x_S$ et $y_M = \gamma y_S$ où γ est le grandissement de la lentille défini par $\gamma = -\frac{d_{det}}{p}$. En effectuant un changement de variable : $x'_M = x_M - \gamma x_S$ et $y'_M = y_M - \gamma y_S$ nous obtenons

$$\begin{aligned} (h_{opt})(x'_M, y'_M) &\propto \left| \iint A(x_P, y_P) e^{-i\phi_{def}(x_P, y_P)} e^{-i\frac{2\pi}{\lambda d_{det}} (x'_M x_P + y'_M y_P)} dx_P dy_P \right|^2 \\ &\propto \left| TF \left(A(x_P, y_P) e^{-i\phi_{def}(x_P, y_P)} \right) \right|^2 \left(\frac{x'_M}{\lambda d_{det}}, \frac{y'_M}{\lambda d_{det}} \right). \end{aligned}$$

La FEP optique h_{opt} est donc calculée par la transformée de Fourier de la fonction pupille multipliée par un terme de phase de défocalisation, centrée sur l'image géométrique du point source.

Génération de scènes vérifiant l'a priori gaussien

Nous présentons ici une méthode de génération de scènes vérifiant l'a priori utilisé dans l'algorithme d'estimation de profondeur présenté dans le chapitre 1 à savoir :

$$p(\mathbf{x}|Q_x) \propto \exp -\frac{\mathbf{x}^t Q_x \mathbf{x}}{2}, \quad (\text{C.1})$$

C.1 Principe général

Soit \mathbf{X} un VIG de taille M correspondant à une fenêtre de la scène de taille $\sqrt{M} \times \sqrt{M}$, vérifiant l'a priori (C.1). Le principe est de construire un VIG $\tilde{\mathbf{X}}$ tel que pour toute matrice A dont les lignes engendrent l'image de Q_x , $A\mathbf{X}$ et $A\tilde{\mathbf{X}}$ sont des vecteurs aléatoires gaussiens propres de même covariance.

Q_x est de taille $M \times M$ et de rang k , nous la décomposons sous la forme :

$$Q_x = V\Delta V^t, \quad (\text{C.2})$$

où V est de dimension $M \times k$ et Δ correspond aux valeurs propres non nulles de Q_x . Soit $\tilde{\mathbf{Y}}_0$ un vecteur aléatoire blanc gaussien de taille k et de covariance égale à l'identité, alors $\tilde{\mathbf{Y}} = \sqrt{\Delta}^{-1} \tilde{\mathbf{Y}}_0$ est un vecteur aléatoire gaussien de covariance Δ^{-1} . Soit $\tilde{\mathbf{X}}$ défini par :

$$\tilde{\mathbf{X}} = V\tilde{\mathbf{Y}} = V\sqrt{\Delta}^{-1} \tilde{\mathbf{Y}}_0, \quad (\text{C.3})$$

les vecteurs $V^t \tilde{\mathbf{X}}$ et $V^t \mathbf{X}$ ont alors la même covariance Δ^{-1} . Il est donc possible de construire $\tilde{\mathbf{X}}$ à partir du vecteur $\tilde{\mathbf{Y}}_0$ et des matrices issues de la diagonalisation de Q_x .

C.2 Pratique

En pratique, nous générons un échantillon du vecteur aléatoire $\tilde{\mathbf{Y}}_0$ à l'aide de la fonction `randn` de Matlab puis nous le multiplions par Δ^{-1} , puis par V , obtenus par décomposition de la matrice Q_x . Ces deux matrices sont notamment obtenues à partir de la diagonalisation de la matrice Q_x construite à partir de la relation définie à la page 26 :

$$Q_x = \frac{D^t D}{\sigma_x^2}, \quad (\text{C.4})$$

où D est la concaténation verticale des matrices de convolution associées aux dérivées horizontales et verticales d'ordre 1. La figure C.1 montre un exemple de scène générée suivant cette méthode, avec une valeur de σ_x^2 de 3.

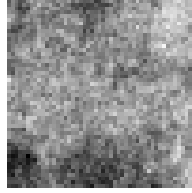


Fig. C.1: Génération d'une fenêtre de scène correspondant à l'a priori gaussien isotrope.

Implémentation du critère GL à partir de la GSVD

Dans le chapitre 1, nous proposons une implémentation du critère GL basée sur la décomposition des matrices de convolution H et D , en utilisant la décomposition en valeurs singulières généralisées (GSVD). L'objectif de cette annexe est de détailler les étapes permettant d'obtenir une expression simple du critère GL. Rappelons que d'après l'équation (1.44) page 28 le critère GL s'écrit :

$$\text{GL}(\psi) = \mathbf{y}^t P_\psi \mathbf{y} |P_\psi|_+^{\frac{-1}{(N-1)}}, \quad (\text{D.1})$$

avec $P_\psi = I - H(H^t H + \alpha D^t D)^{-1} H$.

D.1 Principe

La décomposition GSVD de deux matrices H de dimension $N \times M$ et D de dimension $P \times M$ s'écrit :

$$H = U C X^t \quad (\text{D.2})$$

$$D = V S X^t \quad (\text{D.3})$$

$$I = C^t C + S^t S, \quad (\text{D.4})$$

où U (respectivement V) est une matrice unitaire de taille $N \times N$ (resp. $P \times P$), I est la matrice identité et X une matrice non singulière [Golub et Van Loan, 1996]. Les matrices C et S sont respectivement de taille $N \times r$ et $P \times r$ et , où r est le rang de la concaténation verticale de H et D . Les matrices $C^t C$ et $S^t S$ sont des matrices diagonales de dimension $r \times r$. La structure des matrices C et S dépend de la dimension et du rang des matrices H et D . Dans le cas où $N < r$ et suivant le formalisme utilisé dans **Matlab**, les matrices C et S sont de la forme :

$$S = \begin{bmatrix} \mathbf{I}_{k+l-N} & \mathbf{0}_{k+l-N, N-k} & \mathbf{0}_{k+l-N, k} \\ \mathbf{0}_{N-k, k+l-N} & S_0 & \mathbf{0}_{N-k, k} \\ \mathbf{0}_{P-l, k+l-N} & \mathbf{0}_{P-l, N-k} & \mathbf{0}_{P-l, k} \end{bmatrix} \quad (\text{D.5})$$

$$C = \begin{bmatrix} \mathbf{0}_{N-k, k+l-N} & C_0 & \mathbf{0}_{N-k, k} \\ \mathbf{0}_{k, k+l-N} & \mathbf{0}_{k, N-k} & \mathbf{I}_{k, k} \end{bmatrix}. \quad (\text{D.6})$$

où l est le rang de la matrice D , $k = r - l$ et I_j est la matrice identité de dimension $j \times j$. De plus, $\mathbf{0}_{k_1, k_2}$ est une matrice de taille $k_1 \times k_2$ contenant uniquement des zéros. C_0 et S_0 sont des matrices diagonales de dimension $N - k \times N - k$. Notons que pour notre application, r vaut M , $l=P$ et $k = 1$.

En utilisant les équations (D.2) et (D.3), on peut montrer que la matrice P_ψ peut se mettre sous la forme :

$$P_\psi = U\Omega U^t, \quad (\text{D.7})$$

avec $\Omega = I - C(C^t C + \alpha S^t S)^{-1} C^t$.

D.2 Exemple simple

Pour commencer nous considérons le cas de matrices H et D de dimensions très réduites. Supposons que H soit de taille 3×5 et D de taille 4×5 et de rang 4, soit $M=5$, $N=3$ et $P=4$. Les matrices C et $C^t C$ sont alors respectivement de taille 3×5 et 5×5 et d'après les équations (D.6) et (D.5) elles s'écrivent sous la forme :

$$C = \begin{bmatrix} 0 & 0 & c_1 & 0 & 0 \\ 0 & 0 & 0 & c_2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ et } C^t C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c_1^2 & 0 & 0 \\ 0 & 0 & 0 & c_2^2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (\text{D.8})$$

S est de taille 4×5 et s'écrit sous la forme :

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & s_1 & 0 & 0 \\ 0 & 0 & 0 & s_2 & 0 \end{bmatrix} \text{ et } S^t S = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & s_1^2 & 0 & 0 \\ 0 & 0 & 0 & s_2^2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (\text{D.9})$$

L'équation (D.4) implique que les coefficients c_i et s_i vérifient $c_i^2 + s_i^2 = 1$. Nous pouvons alors en déduire :

$$C^t C + \alpha S^t S = \begin{bmatrix} \alpha & 0 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & 0 \\ 0 & 0 & c_1^2 + \alpha s_1^2 & 0 & 0 \\ 0 & 0 & 0 & c_2^2 + \alpha s_2^2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (\text{D.10})$$

et de plus :

$$C(C^t C + \alpha S^t S)^{-1} C^t = \begin{bmatrix} \frac{c_1^2}{c_1^2 + \alpha s_1^2} & 0 & 0 \\ 0 & \frac{c_2^2}{c_2^2 + \alpha s_2^2} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (\text{D.11})$$

La matrice Ω est donc une matrice diagonale dont les termes sont définis pour i compris entre 1 et 3 par :

$$\Omega(i, i) = \omega_i = \frac{\alpha s_i^2}{c_i^2 + \alpha s_i^2}, \quad (\text{D.12})$$

où l'on a posé $s_3 = 0$ et $c_3 = 1$.

D.3 Expression du critère GL à partir de la GSVD

Nous déduisons de l'exemple précédent que la matrice Ω est une matrice diagonale dont les termes sont définis pour i compris entre 1 et N par :

$$\Omega(i, i) = \frac{\alpha s_i^2}{c_i^2 + \alpha s_i^2}, \quad (\text{D.13})$$

où l'on a posé $s_N = 0$ et $c_N = 1$. Pour calculer le critère GL, nous avons besoin de calculer $|P_\psi|_+$, autrement dit le produit des valeurs propres non nulles de P_ψ qui s'écrit alors :

$$|P_\psi|_+ = |\Omega|_+ = \prod_{i=1}^{N-1} \frac{\alpha s_i^2}{c_i^2 + \alpha s_i^2}. \quad (\text{D.14})$$

De plus en posant $\mathbf{z} = U^t \mathbf{y}$, le terme $\mathbf{y}^t P_\psi \mathbf{y}$ du critère GL s'écrit sous la forme :

$$\mathbf{y}^t U \Omega U^t \mathbf{y} = \|\Omega U^t \mathbf{y}\|^2 = \sum_i^N \frac{\alpha s_i^2}{c_i^2 + \alpha s_i^2} z_i^2. \quad (\text{D.15})$$

Ainsi, en posant $\omega_i = \frac{\alpha s_i^2}{c_i^2 + \alpha s_i^2}$ où \mathbf{c}^2 et \mathbf{s}^2 correspondent respectivement aux N dernières valeurs diagonales de $C^t C$ et $S^t S$ et en posant par convention $s_N = 0$ et $c_N = 1$, nous en déduisons une expression simplifiée du critère GL :

$$\text{GL}(\psi) = \left(\prod_{i=1}^{N-1} \omega_i \right)^{-\frac{1}{N-1}} \left(\sum_{i=1}^N \omega_i z_i^2 \right), \quad (\text{D.16})$$

qui correspond à l'expression indiquée dans l'équation (1.51).

Approximation analytique de la BCR dans le cas mono-image

L'objectif de cette annexe est de préciser les calculs permettant d'approcher la BCR définie dans le chapitre 3 σ_{BCR} de manière analytique, dans le cas SA-DFD où les FEP sont modélisées par une fonction gaussienne et lorsque les matrices de convolution sont diagonalisées dans l'espace de Fourier. Rappelons que dans le chapitre 3, nous avons montré que l'écart-type théorique σ_{BCR} pouvait s'écrire : $\sigma_{\text{BCR}}(p) = A(p)B(p)$ avec :

$$A(p) = \frac{\sqrt{2}p^2 F/\#p_x (p_0 - f)}{8\pi^2 f^2 p_0}, \quad (\text{E.1})$$

$$\simeq \frac{\sqrt{2}p^2 F/\#p_x}{8\pi^2 f^2}, \quad (\text{E.2})$$

et

$$B(p) = \frac{1}{\rho\tau} \frac{1}{\sqrt{\sum_{\mu,\nu} \left((\mu^2 + \nu^2) \frac{\Delta_{\tilde{h}}(\mu,\nu,p)}{\Delta_{\tilde{h}}(\mu,\nu,p) + \alpha\Delta_{\tilde{d}}(\mu,\nu)} \right)^2}}. \quad (\text{E.3})$$

De plus, nous utilisons les approximations suivantes :

$$\Delta_{\tilde{h}}(\mu, \nu, p) \simeq e^{-4\pi^2\tau^2(\mu^2+\nu^2)}, \quad (\text{E.4})$$

$$\Delta_{\tilde{d}}(\mu, \nu) \simeq 4 \left(\sin^2(\pi\mu) + \sin^2(\pi\nu) \right). \quad (\text{E.5})$$

Nous recherchons ici une formule analytique simple pour approcher la fonction B .

Pour commencer, nous posons :

$$S = \sum_{\mu,\nu} \frac{(\mu^2 + \nu^2)^2 \Delta_{\tilde{h}}^2(\mu, \nu, p)}{(\Delta_{\tilde{h}}(\mu, \nu, p) + \alpha\Delta_{\tilde{d}}(\mu, \nu))^2}. \quad (\text{E.6})$$

Le calcul de cette somme S est assimilé au calcul de l'intégrale :

$$I_{nt} = \iint_{\mu\nu} \frac{(\mu^2 + \nu^2)^2 \Delta_{\tilde{h}}^2(\mu, \nu, p)}{(\Delta_{\tilde{h}}(\mu, \nu, p) + \alpha\Delta_{\tilde{d}}(\mu, \nu))^2} N d\mu d\nu \quad (\text{E.7})$$

$$= \iint_{\mu\nu} (\mu^2 + \nu^2)^2 f(\mu, \nu) N d\mu d\nu, \quad (\text{E.8})$$

avec :

$$f(\mu, \nu) = \frac{\Delta_{\tilde{h}}(\mu, \nu, p)^2}{(\Delta_{\tilde{h}}(\mu, \nu, p) + \alpha \Delta_{\tilde{d}}(\mu, \nu))^2}. \quad (\text{E.9})$$

La figure E.1 présente une coupe des fonctions $\alpha\Delta_{\tilde{d}}$, $\Delta_{\tilde{h}}$, $\Delta_{\tilde{d}} + \Delta_{\tilde{h}}$, et f suivant $\nu = 0$. Ces courbes sont tracées pour $\tau = 2$ et $\alpha = 0.001$.

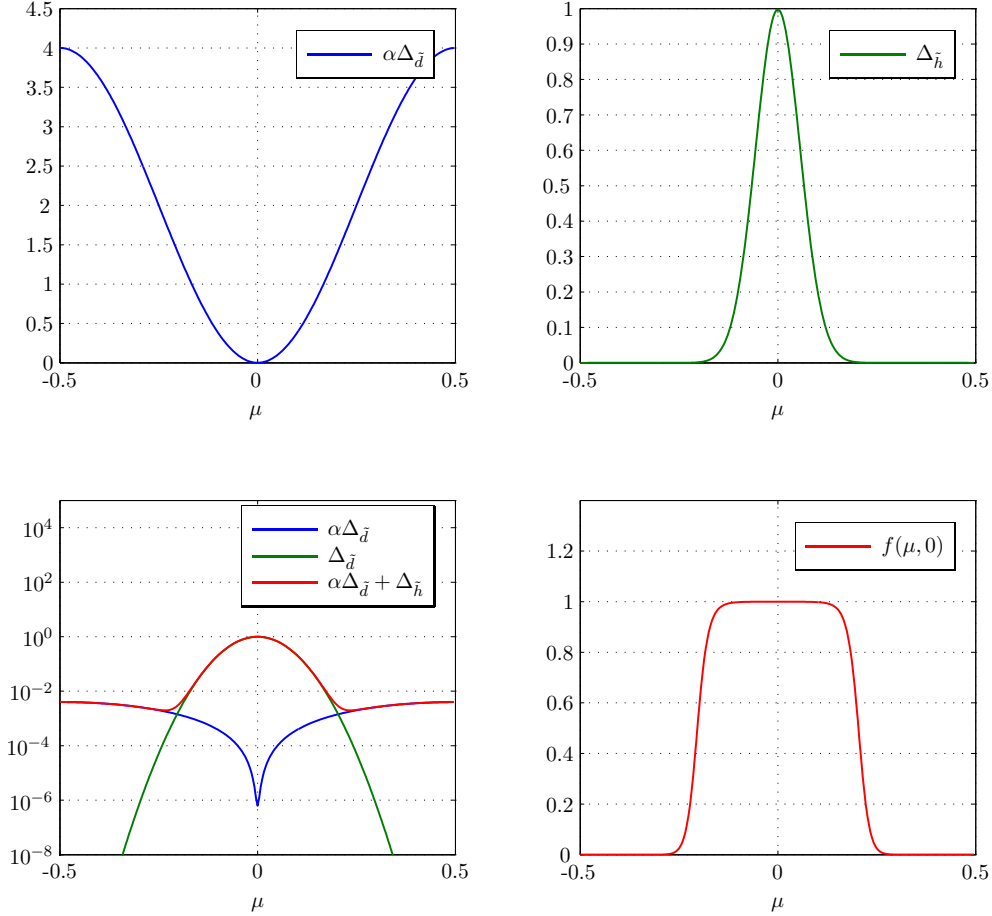


Fig. E.1: Coupes des fonctions $\alpha\Delta_{\tilde{d}}$, $\Delta_{\tilde{h}}$, $\alpha\Delta_{\tilde{d}} + \Delta_{\tilde{h}}$ et f suivant l'axe $\nu = 0$, pour $\tau = 2$ et $\alpha = 0.001$.

Nous pouvons constater que pour les faibles fréquences, la somme $\alpha\Delta_{\tilde{d}} + \Delta_{\tilde{h}}$ est très proche du terme $\Delta_{\tilde{h}}$ car la fonction f est proche de 1. Par ailleurs pour les plus hautes fréquences, la somme $\alpha\Delta_{\tilde{d}} + \Delta_{\tilde{h}}$ suit le terme $\alpha\Delta_{\tilde{d}}$ et f tend vers 0. Les termes les plus élevés de I_{nt} correspondent donc aux fréquences les plus faibles pour lesquelles $f \simeq 1$. Nous considérons alors :

$$I_{nt} \simeq N \int_{-\mu_0-\nu_0}^{\mu_0\nu_0} (\mu^2 + \nu^2)^2 d\mu d\nu. \quad (\text{E.10})$$

où μ_0 et ν_0 correspondent aux fréquences maximales pour lesquelles nous considérons $f = 1$. Nous choisissons de fixer ces fréquences au niveau de l'intersection des courbes de $\Delta_{\tilde{h}}(\mu, \nu)$ et $\alpha\Delta_{\tilde{d}}(\mu, \nu)$, autrement dit les fréquences μ_0 et ν_0 sont solutions de l'équation :

$$4\alpha (\sin^2(\pi\mu_0) + \sin^2(\pi\nu_0)) = e^{-4\pi\tau^2(\mu_0^2 + \nu_0^2)} \quad (\text{E.11})$$

La résolution de l'équation (E.11) n'est pas triviale. C'est pourquoi nous utilisons l'approximation suivante : $4(\sin^2(\pi\mu) + \sin^2(\pi\nu)) \simeq 4\pi^2(\mu^2 + \nu^2)$, approximation valable pour de faibles fréquences. L'équation (E.11) devient alors :

$$4\alpha\pi^2 t^2 = e^{-4\pi\tau^2 t^2}, \quad (\text{E.12})$$

en posant $t^2 = \mu_0^2 + \nu_0^2$. De plus, avec $X = 4\pi\tau^2 t^2$, nous obtenons :

$$Xe^X = \frac{\tau^2}{\alpha}, \quad (\text{E.13})$$

La solution de (E.13) correspond à la fonction W-lambert prise en $\frac{\tau^2}{\alpha}$. Nous pouvons donc écrire $X = W(\frac{\tau^2}{\alpha})$, avec W la fonction W-lambert. Cette fonction existe dans les logiciels de calcul scientifiques tels que **Matlab** et peut donc X peut être déterminé numériquement. Ceci conduit à l'expression de t suivante :

$$t = \frac{1}{2\pi\tau} \sqrt{W\left(\frac{\tau^2}{\alpha}\right)}. \quad (\text{E.14})$$

Notons qu'en considérant $\alpha \ll \tau^2$, il est également possible d'utiliser l'équivalent $W(x) \sim \ln x - \ln \ln x$. Cette approximation revient à considérer que l'objet considéré est éloigné du plan de mise au point.

Finalement, en utilisant les coordonnées polaires, l'équation (E.10) devient :

$$I_{nt} \simeq 2\pi N \int_0^t r^5 dr = \frac{2\pi}{6} N t^6, \quad (\text{E.15})$$

avec t défini par (E.14). Nous obtenons ainsi :

$$I_{nt} \simeq \frac{W(\frac{\tau^2}{\alpha})^3 N}{6(2\pi)^5 \tau^6}. \quad (\text{E.16})$$

En conclusion, en utilisant l'équivalent de la fonction W-lambert, nous obtenons l'approximation de la fonction $B(p)$ suivante :

$$B(p) \sim \frac{\tau^2 (2\pi)^{5/2} \sqrt{6}}{\sqrt{N} \rho (\ln \frac{\tau^2}{\alpha} - \ln \ln \frac{\tau^2}{\alpha})^{3/2}}, \quad (\text{E.17})$$

et en combinant les équations (E.2) et (E.17), σ_{BCR} s'écrit alors :

$$\sigma_{\text{BCR}}(p) \simeq \kappa \frac{p^2 F_{/\#} p_x}{f^2} \frac{\tau^2}{\sqrt{N} \rho (\ln \frac{\tau^2}{\alpha} - \ln \ln \frac{\tau^2}{\alpha})^{3/2}}. \quad (\text{E.18})$$

Avec $\kappa = \sqrt{6\pi}$. Cette expression montre que la variation de l'écart type dépend de la profondeur ainsi que de la taille de la FEP. Les figures E.2 (a) et (b) montrent les valeurs de B et de σ_{BCR} comparées aux approximations proposées dans (E.17) et (E.18), les paramètres optiques utilisés sont présentés dans le tableau E.1. Il existe un écart variant entre moins de 1 et 7% entre les courbes obtenues avec l'équation (E.18) et la valeur numérique calculée par transformée de Fourier. En effet, un certain nombre d'approximations ont été utilisées afin d'obtenir cette formule. Néanmoins l'équation (E.18) est intéressante car elle fait apparaître la profondeur de champ définie dans l'équation (3.5), pondérée par un terme qui dépend de la taille de la FEP et du rapport signal à bruit. Une interprétation plus approfondie de cette formule est réalisée dans le chapitre 3 section 3.3.2.4.

f	25 mm
$F/\#$	2.5
p_x	5 μm
p_0	2 m
Taille des fenêtres	100 \times 100
Taille des FEP	100 \times 100
δ	1 mm

Tableau E.1: Paramètres utilisés pour tracer la figure E.2.

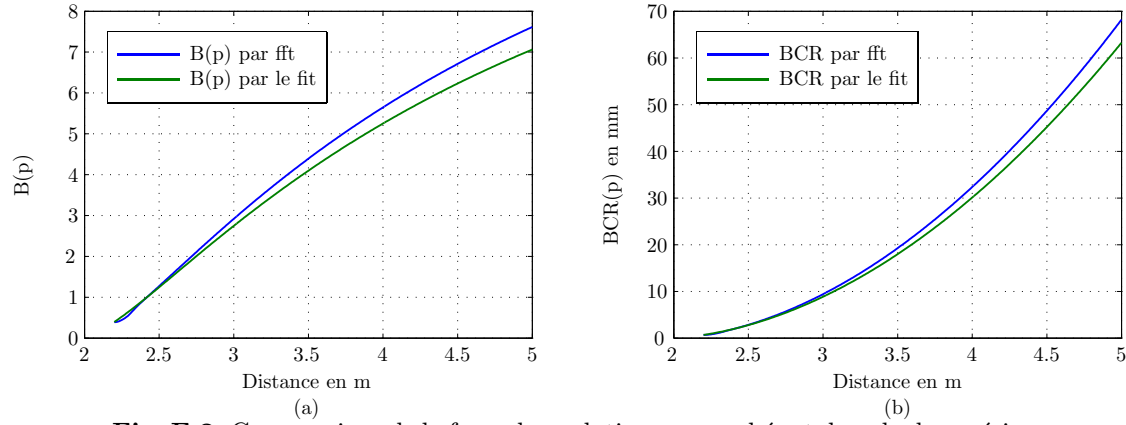


Fig. E.2: Comparaison de la formule analytique approchée et du calcul numérique de B et de σ_{BCR} . (a) Comparaison des valeurs numériques de B calculées à partir de l'approximation (E.17) où par la formule (E.3). (b) Comparaisons des valeurs numériques de σ_{BCR} calculées à partir de l'équation (3.59) ou à partir de l'équation approchée (E.18).

Scènes naturelles



Fig. F.1: Scènes naturelles utilisées pour simuler des images réelles.

Bibliographie

- ABLES, J. (1968). Fourier transform photography : a new method for x-ray astronomy. *In Proc. of the Astronomical Society of Australia*, volume 1, pages 172–174. Cité page [54](#)
- ADELSON, E. et WANG, J. (1992). Single lens stereo with a plenoptic camera. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 14, pages 99–106. Cité page [3](#)
- AGUET, F., VAN DE VILLE, D. et UNSER, M. (2005). A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles. *Optics Express*, volume 13, pages 10503–10522. Cité pages [99](#) et [108](#)
- ALLEYSSON, D., SÜSTRUNK, S. et HÉRAULT, J. (2005). Linear demosaicing inspired by the human visual system. *IEEE Trans. on Image Processing (IP)*, volume 14, pages 439–457. Cité page [143](#)
- BAE, S. et DURAND, F. (2007). Defocus magnification. *Computer Graphics Forum*, volume 26, pages 571–579. Cité pages [xi](#), [20](#) et [21](#)
- BANDO, Y., CHEN, B. et NISHITA, T. (2008). Extracting depth and matte using a color-filtered aperture. *ACM Trans. on Graphics (TOG)*, volume 27, pages 1–9. Cité page [59](#)
- BESAG, J. et KOOPERBERG, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, volume 82, page 733. Cité pages [27](#), [102](#), [103](#) et [188](#)
- BISHOP, T. et FAVARO, P. (2011). The light field camera : Extended depth of field, aliasing and super-resolution. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 34(5), pages 972–986. Cité page [3](#)
- BLANC, A. (2002). *Identification de réponses impulsionnelles et restauration d’images : apport de la diversité de phase*. Thèse de doctorat, Université Paris XI Orsay. Cité page [22](#)
- BLANCO, L. et MUGNIER, L. (2011). Marginal blind deconvolution of adaptive optics retinal images. *Optics Express*, volume 19, pages 23227–23239. Cité page [21](#)
- BLAYVAS, I., KIMMEL, R. et RIVLIN, E. (2007). Role of optics in the accuracy of depth-from-defocus systems. *Journal of the Optical Society of America (A)*, volume 24, pages 967–972. Cité page [97](#)
- BLENDOWSKE, R. (2007). Role of optics in the accuracy of depth-from-defocus systems : comment. *Journal of the Optical Society of America (A)*, volume 24, pages 3242–3244. Cité page [97](#)

- BOVE JR, V. *et al.* (1993). Entropy-based depth from focus. *Journal of the Optical Society of America (A)*, volume 10, pages 561–566. Cité page [16](#)
- BRADY, D., DOGARIU, A., FIDDY, M. *et* MAHALANOBIS, A. (2008). Computational optical sensing and imaging : introduction to the feature issue. *Applied Optics*, volume 47, pages COSI1–COSI2. Cité page [5](#)
- CAI, J., JI, H., LIU, C. *et* SHEN, Z. (2009). Blind motion deblurring from a single image using sparse approximation. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 104–111. Cité page [19](#)
- CAO, Y., FANG, S. *et* WANG, F. (2011). Single image multi-focusing based on local blur estimation. In *Int. Conf. on Image and Graphics (ICIG)*, pages 168 –175. Cité pages [20](#) *et* [42](#)
- CATHEY, W. *et* DOWSKI, E. (2002). New paradigm for imaging systems. *Applied Optics*, volume 41, pages 6080–6092. Cité page [6](#)
- CHAKRABARTI, A. *et* ZICKLER, T. (2012). Depth and deblurring from a spectrally varying depth of field. In *IEEE Europ. Conf. Comp. Vision (ECCV)*, volume Part V. Cité pages [59](#) *et* [182](#)
- CHAKRABARTI, A., ZICKLER, T. *et* FREEMAN, W. (2010). Analyzing spatially-varying blur. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 2512–2519. Cité pages [xii](#), [14](#), [23](#), [39](#), [40](#), [41](#) *et* [182](#)
- CHAMPAGNAT, F. (2012). Inference with gaussian improper distributions. Rapport technique RT 5/14983 DTIM. Cité pages [24](#), [102](#) *et* [187](#)
- CHAMPAGNAT, F., DRUART, G., LE BESNERAIS, G. *et* GUÉRINEAU, N. (2010). Concepts innovants en optique non conventionnelle *et* co-conception. Rapport technique, PRF SPIDER Lot 1.RT 1/14983 DTIM/DOTA. Cité page [152](#)
- CHAMPAGNAT, F., LE BESNERAIS, G. *et* KULCSÁR, C. (2009). Statistical performance modeling for superresolution : a discrete data-continuous reconstruction framework. *Journal of the Optical Society of America (A)*, volume 26, pages 1730–1746. Cité page [6](#)
- CHANAS, L., TARCHOUNA, I., GUICHARD, F. *et* LIÈGE, B. (2005). Brevet. Procédé d’amélioration de la netteté d’au moins une couleur d’une image numérique. (FR2880958). Cité page [59](#)
- CHO, S. *et* LEE, S. (2009). Fast motion deblurring. *ACM Trans. on Graphics (TOG)*, volume 28, pages 145 :1–145–8. Cité page [19](#)
- CHO, T., LEVIN, A., DURAND, F. *et* FREEMAN, W. (2010). Motion blur removal with orthogonal parabolic exposures. In *IEEE Int. Conf. on Comp. Photography (ICCP)*, pages 1–8. Cité page [6](#)
- CHOW, J., ANG, K., LICHTI, D. *et* TESKEY, W. (2012). Performance analysis of a low-cost triangulation-based 3d camera : Microsoft kinect system. In *Int. Soc. for Photogrammetry and Remote Sensing Congress (ISPRS)*, volume XXXIX-B5, pages 175–180. Cité page [63](#)
- CIERPKA, C., ROSSI, M., SEGURA, R. *et* KÄHLER, C. (2010). On the calibration of astigmatism particle tracking velocimetry for microflows. *Measurement Science and Technology*, volume 22, pages 15401–15410. Cité page [184](#)
- CONDAT, L. (2009). A generic variational approach for demosaicking from an arbitrary color filter array. In *IEEE Int. Conf. on Image Processing (ICIP)*, pages 1625–1628. Cité pages [78](#), [80](#), [83](#), [134](#) *et* [183](#)

- COSSAIRT, O. (2010). Spectral focal sweep : Extended depth of field from chromatic aberrations. In *IEEE Int. Conf. on Comp. Photography (ICCP)*, pages 1–8. Cité pages [7](#) et [59](#)
- DELBRACIO, M., MUSÉ, P., ALMANSA, A. et MOREL, J. (2011). The non-parametric sub-pixel local point spread function estimation is a well posed problem. *Int. Journal of Computer Vision (IJCV)*, pages 1–20. Cité pages [vii](#), [xiii](#), [xiv](#), [xviii](#), [60](#), [61](#), [67](#), [68](#), [85](#) et [169](#)
- DIAZ, F. (2011). *Système d'imagerie hybride par codage de pupille*. Thèse de doctorat, Université Paris Sud XI. Cité page [7](#)
- DIAZ, F., GOUDAIL, F., LOISEAUX, B. et HUIGNARD, J. (2009). Design of a complex filter for depth of focus extension. *Optics Letters*, volume 34, pages 1171–1173. Cité page [6](#)
- DICKE, R. (1968). Scatter-hole cameras for x-rays and gamma rays. *The Astrophysical journal*, volume 153, pages L101–L106. Cité page [54](#)
- DOWSKI JR, E. et CATHEY, W. (1994). Single-lens single-image incoherent passive-ranging systems. *Applied Optics*, volume 33, pages 6762–6773. Cité page [7](#)
- ELDER, J. et ZUCKER, S. (1998). Local scale control for edge detection and blur estimation. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 20, pages 699–716. Cité pages [xi](#) et [20](#)
- ENS, J. et LAWRENCE, P. (1993). An investigation of methods for determining depth from focus. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 15, pages 97–108. Cité page [17](#)
- FARID, H. et SIMONCELLI, E. (1998). Range estimation by optical differentiation. *Journal of the Optical Society of America (A)*, volume 15, pages 1777–1786. Cité page [55](#)
- FAVARO, P., BURGER, M. et SOATTO, S. (2004). Scene and motion reconstruction from defocused and motion-blurred images via anisotropic diffusion. In *IEEE Europ. Conf. Comp. Vision (ECCV)*, pages 257–269. Cité pages [17](#) et [18](#)
- FAVARO, P. et SOATTO, S. (2005). A geometric approach to shape from defocus. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 27, pages 406–417. Cité pages [18](#) et [23](#)
- FAVARO, P., SOATTO, S., BURGER, M. et OSHER, S. (2008). Shape from defocus via diffusion. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 30, pages 518–531. Cité page [17](#)
- FELZENSZWALB, P. et HUTTENLOCHER, D. (2004). Efficient graph-based image segmentation. *Int. Journal of Computer Vision (IJCV)*, volume 59, pages 167–181. Cité page [182](#)
- FERGUS, R., SINGH, B., HERTZMANN, A., ROWEIS, S. et FREEMAN, W. (2006). Removing camera shake from a single photograph. *ACM SIGGRAPH*, pages 787–794. Cité pages [19](#), [20](#), [21](#) et [33](#)
- FOURNIER, C., DENIS, L. et FOURNEL, T. (2010). On the single point resolution of on-axis digital holography. *Journal of the Optical Society of America (A)*, volume 27, pages 1856–1862. Cité page [99](#)
- GALATSANOS, N., WERNICK, M., KATSAGGELOS, A. et MOLINA, R. (2000). Multichannel image recovery. *Handbook of Image and Video Processing*, volume 12, pages 155–168. Cité page [183](#)

- GARCIA, J., SÁNCHEZ, J., ORRIOLS, X. et BINEFA, X. (2000). Chromatic aberration and depth extraction. In *IEEE Int. Conf. on Pattern Recogn. (ICPR)*, volume 1, pages 762–765. Cité page [59](#)
- GEORGIEV, T. (2010). Rich image capture with plenoptic cameras. In *IEEE Int. Conf. on Comp. Photography (ICCP)*, pages 1–8. Cité page [3](#)
- GOKSTORP, M. (1994). Computing depth from out-of-focus blur using a local frequency representation. In *IEEE Int. Conf. on Pattern Recogn. (ICPR)*, volume 1, pages 153–158. Cité page [16](#)
- GOLUB, G., HEATH, M. et WAHBA, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, volume 21, pages 215–223. Cité page [34](#)
- GOLUB, G. et VAN LOAN, C. (1996). *Matrix computations*, volume 3. Johns Hopkins University Press. Cité page [197](#)
- GOODMAN, J. (1996). *Introduction to Fourier Optics*. McGraw-Hill, 2nd édition. Cité page [193](#)
- GOTOH, T. et OKUTOMI, M. (2004). Direct super-resolution and registration using raw CFA images. volume 2, pages 600–607. Cité page [183](#)
- GOTTESMAN, S. et FENIMORE, E. (1989). New family of binary arrays for coded aperture imaging. *Applied Optics*, volume 28, pages 4344–4352. Cité pages [xiii](#), [54](#) et [56](#)
- GREEN, P., SUN, W., MATUSIK, W. et DURAND, F. (2007). Multi-aperture photography. *ACM Trans. on Graphics (TOG)*, volume 26, pages 68 :1–68 :7. Cité pages [19](#) et [58](#)
- GREENGARD, A., SCHECHNER, Y. et PIESTUN, R. (2006). Depth from diffracted rotation. *Optics Express*, volume 31, pages 181–183. Cité pages [99](#) et [184](#)
- GUICHARD, F., NGUYEN, H., TESSIÈRES, R., PYANET, M., TARCHOUNA, I. et CAO, F. (2009). Extended depth-of-field using sharpness transport across color channels. In *Proc. of SPIE*, volume 7250. Cité pages [7](#), [59](#), [84](#) et [91](#)
- HALMOS, P. (1969). *Measure theory*. Int. Student Editions. D. Van Nostrand Company, New York, N.Y. Cité page [187](#)
- HASINOFF, S. et KUTULAKOS, K. (2008). Light-efficient photography. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 33, pages 2203–2214. Cité page [6](#)
- HIRSCHMULLER, H. et SCHARSTEIN, D. (2009). Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 31, pages 1582–1599. Cité page [3](#)
- HIURA, S. et MATSUYAMA, T. (1998). Depth measurement by the multi-focus camera. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 953–959. Cité pages [xiii](#), [19](#), [54](#), [56](#) et [58](#)
- HU, H. et DE HAAN, G. (2007). Adaptive image restoration based on local robust blur estimation. In *Proc. of the Int. Conf. on Advanced concepts for intelligent vision*, pages 461–472. Cité pages [20](#) et [42](#)
- HUNT, B. R. (1973). The application of constrained least squares estimation to image restoration by digital computer. *IEEE Trans. on Computers*, volume C-22, pages 805–812. Cité page [106](#)
- HWANG, T., CLARK, J. et YUILLE, A. (1989). A depth recovery algorithm using defocus information. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 476–482. Cité page [16](#)

- IDIER, J., éditeur (2001). *Approche bayésienne pour les problèmes inverses*. Traité IC2, Série traitement du signal et de l'image, Hermès, Paris. Cité pages [25](#), [28](#), [106](#), [126](#) et [132](#)
- JOHNSON, G., DOWSKI, E. et CATHEY, W. (2000). Passive ranging through wave-front coding : information and application. *Applied Optics*, volume 39, pages 1700–1710. Cité pages [7](#) et [99](#)
- JOSHI, N., KANG, S., ZITNICK, C. et SZELISKI, R. (2010). Image deblurring using inertial measurement sensors. *ACM Trans. on Graphics (TOG)*, volume 29, pages 30 :1–30 :9. Cité page [6](#)
- JOSHI, N., SZELISKI, R. et KRIEGMAN, D. (2008). PSF estimation using sharp edge prediction. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 1–8. Cité pages [21](#) et [61](#)
- KAY, S. (1993). *Fundamentals of statistical signal processing : estimation theory*. Prentice-Hall. Cité page [98](#)
- KLARQUIST, W., GEISLER, W. et BOVIK, A. (1995). Maximum-likelihood depth-from-defocus for active vision. In *nt. Conf. on Intelligent Robots and Systems*, volume 3, pages 374–379. Cité page [23](#)
- KOLMOGOROV, V. et ZABIH, R. (2004). What energy functions can be minimized via graph cuts? *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 26, pages 147–159. Cité page [181](#)
- KUNDUR, D. et HATZINAKOS, D. (1996). Blind image deconvolution. *IEEE Signal Processing Magazine*, volume 13, pages 43–64. Cité pages [19](#) et [20](#)
- LEVIN, A. (2007). Blind motion deblurring using image statistics. *Advances in Neural Information Processing Systems*, volume 19, pages 841–848. Cité pages [14](#) et [23](#)
- LEVIN, A., FERGUS, R., DURAND, F. et FREEMAN, W. (2007a). Deconvolution using natural image priors. *ACM Trans. on Graphics (TOG)*, volume 26, pages 0–2. Cité page [25](#)
- LEVIN, A., FERGUS, R., DURAND, F. et FREEMAN, W. (2007b). Image and depth from a conventional camera with a coded aperture. *ACM Trans. on Graphics (TOG)*, volume 26, pages 70 :1–70 :9. Cité pages [vii](#), [xi](#), [xii](#), [xiii](#), [5](#), [7](#), [8](#), [9](#), [10](#), [15](#), [21](#), [22](#), [25](#), [45](#), [46](#), [47](#), [48](#), [53](#), [55](#), [56](#), [60](#), [61](#), [66](#), [67](#), [93](#), [96](#), [122](#), [181](#) et [182](#)
- LEVIN, A., SAND, P., CHO, T., DURAND, F. et FREEMAN, W. (2008). Motion-invariant photography. *ACM Trans. on Graphics (TOG)*, volume 27, pages 71 :1–71 :9. Cité page [6](#)
- LEVIN, A., WEISS, Y., DURAND, F. et FREEMAN, W. (2009). Understanding and evaluating blind deconvolution algorithms. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 1–8. Cité pages [19](#), [20](#), [21](#), [25](#) et [33](#)
- LEVIN, A., WEISS, Y., DURAND, F. et FREEMAN, W. (2011). Efficient marginal likelihood optimization in blind deconvolution. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 2657–2664. Cité pages [19](#) et [21](#)
- LI, F., YU, J. et CHAI, J. (2008). A hybrid camera for motion deblurring and depth map super-resolution. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 1–8. Cité page [6](#)
- LIM, J., KANG, J. et OK, H. (2008). Robust local restoration of space-variant blur image. In *Proceedings of SPIE*, volume 6817. Cité page [59](#)
- LÉTIENNE, A. (2012). *Super-Résolution : Développement d'algorithmes rapides et évaluation de performance*. Thèse de doctorat, Université Paris 13-Institut Galilée. Cité page [6](#)

- MAIK, V., CHO, D., KIM, S., HAR, D. et PAIK, J. (2010). Computational filter-aperture approach for single-view multi-focusing. *In IEEE Int. Conf. on Image Processing (ICIP)*, pages 1541–1544. Cité page 59
- MARTINELLO, M., BISHOP, T. et FAVARO, P. (2010). A Bayesian approach to shape from coded aperture. *In IEEE Int. Conf. on Image Processing (ICIP)*, pages 3521–3524. Cité pages 5, 7, 21, 23 et 182
- MARTINELLO, M. et FAVARO, P. (2011a). Fragmented aperture imaging for motion and defocus deblurring. *In IEEE Int. Conf. on Image Processing (ICIP)*, pages 3413–3416. Cité page 23
- MARTINELLO, M. et FAVARO, P. (2011b). Single image blind deconvolution with higher-order texture statistics. *Video Processing and Computational Video*, pages 124–151. Cité pages xii, 5, 7, 8, 21, 23, 45, 46, 47, 48, 55, 66, 96 et 182
- MILGROM, B., KONFORTI, N., GOLUB, M. et MAROM, E. (2010). Novel approach for extending the depth of field of Barcode decoders by using RGB channels of information. *Optics Express*, volume 18, pages 17027–17039. Cité page 59
- MORRISON, J. (2011). *Vision aided inertial navigation system augmented with a coded aperture*. Thèse de doctorat, Airforce Institute of Technology. Cité pages xiii, 55 et 56
- NAGAHARA, H., KUTHIRUMMAL, S., ZHOU, C. et NAYAR, S. (2008). Flexible depth of field photography. *In IEEE Europ. Conf. Comp. Vision (ECCV)*, pages 60–73. Cité page 7
- NAGAHARA, H., ZHOU, C., WATANABE, T., ISHIGURO, H. et NAYAR, S. (2010). Programmable aperture camera using lcos. *In IEEE Europ. Conf. Comp. Vision (ECCV)*, pages 337–350. Cité pages 19 et 58
- NAIR, H. et STEWART, C. (1992). Robust focus ranging. *In IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 309–314. Cité page 4
- NAMBOODIRI, V. et CHAUDHURI, S. (2008). Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera. *In IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 1–6. Cité page 21
- NAYAR, S. (1992). Shape from focus system. *In IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 302–308. Cité page 4
- NAYAR, S. (2011). Computational camera : Approaches, benefits and limits. Rapport technique. Cité page 6
- NAYAR, S. et BEN-EZRA, M. (2004). Motion-based motion deblurring. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 26, pages 689–698. Cité page 6
- NEUMAIER, A. (1998). Solving ill-conditioned and singular linear systems : A tutorial on regularization. *SIAM Rev.*, volume 40, pages 636–666, Philadelphia, PA, USA. Cité pages 29, 34 et 187
- NG, R., LEVOY, M., BRÉDIF, M., DUVAL, G., HOROWITZ, M. et HANRAHAN, P. (2005). Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, volume 2. Cité page 3
- ORIEUX, F., GIOVANNELLI, J. et RODET, T. (2010). Bayesian estimation of regularization and point spread function parameters for wiener-hunt deconvolution. *Journal of the Optical Society of America (A)*, volume 27, pages 1593–1607. Cité page 27
- PAVANI, S., THOMPSON, M., BITEEN, J., LORD, S., LIU, N., TWIEG, R., PIESTUN, R. et MOERNER, W. (2009). Three-dimensional, single-molecule fluorescence imaging beyond

- the diffraction limit by using a double-helix point spread function. *In Proc. of the National Academy of Sciences*, volume 106, pages 2995–2999. Cité page [184](#)
- PENTLAND, A. (1987). A new sense for depth of field. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 4, pages 523–531. Cité pages [5](#) et [16](#)
- RAJAGOPALAN, A. et CHAUDHURI, S. (1997a). Optimal selection of camera parameters for recovery of depth from defocused images. *In IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 219–224. Cité pages [17](#) et [99](#)
- RAJAGOPALAN, A. et CHAUDHURI, S. (1997b). Space-variant approaches to recovery of depth from defocused images. *Computer Vision and Image Understanding*, volume 68, pages 309–329. Cité pages [5](#) et [16](#)
- RAJAGOPALAN, A. et CHAUDHURI, S. (1998). Performance analysis of maximum likelihood estimator for recovery of depth from defocused images and optimal selection of camera parameters. *In Int. Journal of Computer Vision (IJCV)*, volume 30, pages 175–190. Cité pages [xvi](#), [17](#), [99](#), [105](#), [128](#), [129](#), [130](#) et [157](#)
- RAJAGOPALAN, A. et CHAUDHURI, S. (1999). An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 21, pages 577–589. Cité page [17](#)
- RAJAGOPALAN, A., CHAUDHURI, S. et CHELLAPPA, R. (2000). Quantitative analysis of error bounds in the recovery of depth from defocused images. *Journal of the Optical Society of America (A)*, volume 17, pages 1722–1731. Cité pages [17](#) et [99](#)
- RAJAN, D. et CHAUDHURI, S. (2003). Simultaneous estimation of super-resolved scene and depth map from low resolution defocused observations. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 25, pages 1102–1117. Cité page [17](#)
- REEVES, S. et MERSEREAU, R. (1992). Blur identification by the method of generalized cross-validation. *IEEE Trans. on Image Processing (IP)*, volume 1, pages 301–311. Cité pages [19](#) et [34](#)
- ROBINSON, M., STORK, D. et INNOVATIONS, R. (2008). Joint digital-optical design of imaging systems for grayscale objects. *In Proc. of SPIE*, numéro 7100, pages 710011.1–710011.9. Cité page [59](#)
- SCHECHNER, Y. et KIRYATI, N. (1999). The optimal axial interval in estimating depth from defocus. *In IEEE Int. Conf. Comp. Vision (ICCV)*, volume 2, pages 843–848. Cité page [98](#)
- SCHECHNER, Y. et KIRYATI, N. (2000). Depth from defocus vs. stereo : How different really are they? *Int. Journal of Computer Vision (IJCV)*, volume 39, pages 141–162. Cité page [98](#)
- SHAN, Q., JIA, J. et AGARWALA, A. (2008). High-quality motion deblurring from a single image. *ACM Trans. on Graphics (TOG)*, volume 27, pages 73 :1–73 :10. Cité page [19](#)
- SOATTO, S. et FAVARO, P. (2000). A geometric approach to blind deconvolution with application to shape from defocus. *In IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, volume 2, pages 10–17. Cité page [18](#)
- SOREL, M. et SROUBEK, F. (2009). Space-variant deblurring using one blurred and one underexposed image. *IEEE Trans. on Image Processing (IP)*, pages 157–160. Cité page [6](#)
- SOULEZ, F. et THIÉBAUT, E. (2009). Joint deconvolution and demosaicing. *IEEE Trans. on Image Processing (IP)*, pages 145–148. Cité pages [92](#) et [153](#)

- STORK, D. et ROBINSON, M. (2008). Theoretical foundations for joint digital-optical analysis of electro-optical imaging systems. *Applied Optics*, volume 47, pages 64–75. Cité page 7
- SUBBARAO, M. (1988). Parallel depth recovery by changing camera parameters. In *IEEE Int. Conf. Comp. Vision (ICCV)*, pages 149–155. Cité pages 5, 16 et 17
- SUBBARAO, M. et CHOI, T. (1995). Accurate recovery of three-dimensional shape from image focus. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 17, pages 266–274. Cité page 4
- SUBBARAO, M. et SURYA, G. (1993). Application of spatial-domain convolution/deconvolution transform for determining distance from image defocus. In *Proc. of SPIE*, volume 159, pages 159–167. Cité page 17
- SUBBARAO, M. et SURYA, G. (1994). Depth from defocus : a spatial domain approach. *Int. Journal of Computer Vision (IJCV)*, volume 13, pages 271–294. Cité pages 5 et 17
- TAI, Y. et BROWN, M. S. (2009). Single image defocus map estimation using local contrast prior. In *IEEE Int. Conf. on Image Processing (ICIP)*, pages 1797 – 1800. Cité pages 20 et 42
- TAI, Y., BROWN, M. S. et LIN, S. (2009). Detail recovery for single image defocus blur. *IPSP Trans. on Computer Vision and Applications*, volume 1, pages 95–104. Cité pages 20 et 42
- TAI, Y., DU, H., BROWN, M. et LIN, S. (2008). Image/video deblurring using a hybrid camera. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 1–8. Cité page 6
- TEKALP, R. et BIEMOND, J. (1990). Maximum likelihood image and blur identification : a unifying approach. *Optical Engineering*, volume 29, pages 422–435. Cité page 19
- TORREAO, J. et FERNANDES, J. (2005). Single-image shape from defocus. In *Symp. on Computer Graphics and Image Processing*, pages 241–246. Cité pages 20 et 42
- VEERARAGHAVAN, A., RASKAR, R., AGRAWAL, A., MOHAN, A. et TUMBLIN, J. (2007). Dappled photography : Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. on Graphics (TOG)*, volume 26, page 69. Cité pages xiii, 5, 7, 21, 22, 55 et 56
- WAHBA, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, volume 13, pages 1378–1402. Cité pages 28, 34 et 187
- WANG, J., LI, J., GRAY, R. et WIEDERHOLD, G. (2001). Unsupervised multiresolution segmentation for images with low depth of field. *IEEE Trans. on Pattern Anal. Mach. Intell. (PAMI)*, volume 23, pages 85–90. Cité page 181
- WATANABE, M. et NAYAR, S. (1998). Rational filters for passive depth from defocus. *Int. Journal of Computer Vision (IJCV)*, volume 27, pages 203–225. Cité page 16
- XIONG, Y. et SHAFER, S. (1993). Depth from focusing and defocusing. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 68–73. Cité page 4
- XU, L. et JIA, J. (2010). Two-phase kernel estimation for robust motion deblurring. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 157–170. Cité page 19
- YUAN, L., SUN, J., QUAN, L. et SHUM, H. (2007). Image deblurring with blurred/noisy image pairs. *ACM Trans. on Graphics (TOG)*, volume 26, pages 1–9. Cité page 6

- ZHANG, W. et CHAM, W. (2009). Single image focus editing. *Workshop of IEEE Int. Conf. Comp. Vision (ICCV)*, pages 1947–1954. Cité page [21](#)
- ZHOU, C. et NAYAR, S. (2009a). Coded aperture pairs for depth from defocus. In *IEEE Int. Conf. Comp. Vision (ICCV)*, pages 325–332. Cité pages [xiii](#), [5](#), [7](#), [21](#), [22](#), [55](#), [56](#) et [96](#)
- ZHOU, C. et NAYAR, S. (2009b). What are good apertures for defocus deblurring? In *IEEE Int. Conf. on Comp. Photography (ICCP)*, pages 1–8. Cité pages [xiii](#), [7](#), [55](#) et [56](#)
- ZHOU, C. et NAYAR, S. (2011). Computational cameras : Convergence of optics and processing. *IEEE Trans. on Image Processing (IP)*, numéro 99, pages 1–1. Cité page [6](#)
- ZHUO, S., GUO, D. et SIM, T. (2010). Robust flash deblurring. In *IEEE Conf. Comp. Vision Pattern Recogn. (CVPR)*, pages 2440–2447. Cité page [6](#)
- ZHUO, S. et SIM, T. (2009). On the recovery of depth from a single defocused image. In *Computer Analysis of Images and Patterns*, volume 5702, pages 889–897. Springer. Cité pages [xi](#), [20](#), [42](#) et [43](#)
- ZHUO, S. et SIM, T. (2011). Defocus map estimation from a single image. *Pattern Recognition*, volume 44, pages 1852–1858. Cité pages [xi](#), [xii](#), [5](#), [20](#), [21](#), [42](#), [43](#), [44](#) et [45](#)

Publications et communications scientifiques

Publications dans des revues d'audience internationale

P. Trouvé, F. Champagnat, G. Le Besnerais et J. Idier. Unsupervised single image depth from defocus, *soumis à IEEE Transactions on image processing (IP)*.

P. Trouvé, F. Champagnat, J. Sabater, T. Avignon, G. Druart, G. Le Besnerais et J. Idier. Depth estimation using chromatic depth from defocus, a new algorithm and experimental validation, *rédigé et à soumettre dans Applied Optics*.

Brevet

Dépôt d'une enveloppe Soleau : *Procédé de conception d'un imageur monovoie passif capable d'estimer la profondeur*. P. Trouvé, F. Champagnat, G. Le Besnerais et G. Druart, le 8 mars 2011.

Dépôt d'une demande de brevet n° FR 12.02774 : *Procédé de conception d'un imageur monovoie passif capable d'estimer la profondeur*. P. Trouvé, F. Champagnat, G. Le Besnerais et G. Druart, le 17 octobre 2012.

Communications lors de congrès internationaux à comité de sélection et actes publiés

P. Trouvé, F. Champagnat, G. Le Besnerais et J. Idier. Single image local blur identification, *IEEE Int. Conf. on Image Processing (ICIP)*, pages 613 - 616, Septembre 2011, (présentation orale).

P. Trouvé, F. Champagnat, G. Le Besnerais, G. Druart et J. Idier. Chromatic depth from defocus, a theoretical and experimental performance study, *Computational Optical Sensing and Imaging (COSI)*, Juin 2012 (présentation orale).

Communications lors de congrès nationaux sans acte

P. Trouvé, F. Champagnat, G. Le Besnerais et J. Idier. Identification locale du flou à partir d'une seule image. *GDR isis et SOC/SIP*, Mars 2011.

P. Trouvé, F. Champagnat, G. Le Besnerais et J. Idier. Estimation de profondeur à partir d'un imageur monovoie et d'un traitement d'estimation locale de PSF. *Journées d'Imagerie Optique non Conventionnelle (JIONC)*, Avril 2011.

P. Trouvé, F. Champagnat, J. Sabater, T. Avignon, G. Druart, G. Le Besnerais et J. Idier. Imageur chromatique à capacité 3D : première réalisation expérimentale et caractérisation de performance. *Journées d'Imagerie Optique non Conventionnelle (JIONC)*, Mars 2012.

RESUMÉ

CONCEPTION CONJOINTE OPTIQUE/TRAITEMENT POUR UN IMAGEUR COMPACT À CAPACITÉ 3D

Cette thèse traite de la conception d'un imageur monovoie passif capable de produire une carte de profondeur de la scène observée. Cet imageur est conçu en optimisant conjointement les paramètres de l'optique et des traitements, méthode désignée par le terme de *co-conception* ou de *conception conjointe*. La capacité 3D de cet imageur repose sur le concept de *depth from defocus* (DFD) qui consiste à estimer la profondeur à l'aide d'une estimation locale du flou de défocalisation. Ces travaux portent en premier lieu sur le développement d'un algorithme d'estimation locale de profondeur non supervisé et applicable à une famille étendue d'imageurs passifs monovoies. Puis deux concepts d'optique favorisant l'estimation de profondeur sont étudiés, du point de vue théorique et expérimental : l'utilisation d'une pupille codée ou d'une optique avec une aberration chromatique longitudinale non corrigée. La réalisation d'un imageur chromatique à capacité 3D est innovante et permet d'illustrer les avantages de cette solution en termes de précision et de région de l'espace où l'estimation de profondeur est possible. Un modèle de performance est ensuite proposé pour prédire la précision des imageurs utilisant la DFD en fonction des paramètres de l'optique, du capteur et des traitements. Ce modèle est utilisé pour la conception du premier imageur chromatique à capacité 3D co-conçu dont la réalisation a permis d'illustrer l'intérêt de la co-conception.

Mots-clefs : Conception conjointe, estimation de profondeur, flou de défocalisation, modèle de performance, chromatisme.

ABSTRACT

DIGITAL-OPTICAL CODESIGN OF AN IMAGING SYSTEM WITH 3D ABILITY

This thesis is dedicated to the design of a single-frame passive imaging system able to produce a depth map of the observed scene. This design is based on the joint optimisation of the optical and processing parameters, also it is referred to as *codesign*. The depth estimation ability of a single-frame passive imaging system relies on the *depth from defocus* principle (DFD) where depth is estimated based on the local estimation of the defocus blur. A new local depth estimation algorithm is developed. It is unsupervised and generic enough to be used with a large family of imaging systems. Then two optical concepts improving depth estimation accuracy are theoretically and experimentally investigated : a coded aperture or a lens with longitudinal chromatic aberration. The realisation of a chromatic imaging system with depth estimation ability is new and this work underlines the advantages of the chromatic solution, in terms of accuracy and range where depth can be estimated. Then a performance model is developed in order to predict the theoretical depth estimation accuracy of any imaging system that uses DFD. This model predicts the accuracy of an imaging system based on the knowledge of the optical and the processing parameters. It is then used as a tool for the design of the first codesigned chromatic imaging system optimised for depth estimation. The realisation of this prototype has highlighted the interest of the codesign approach.

Keywords : Codesign, depth from defocus, performance model, chromatic aberration.

Discipline : Sciences de l'Ingénieur